This manuscript studied the changes in statistical distributions of sub-daily surface temperatures, dewpoint temperatures, as well as wind speeds, using station-based HadISD dataset. Both zonally averaged quantities and the spatial distributions were considered, and a quantile regression analysis was also performed. Besides the changes of the mean values, different statistical moments were also studied. This work provided great details about the changes of the temperatures and wind speed, in context of global warming. Roughly speaking, I think this manuscript can be a good reference for people who studies the effects of global warming. However, to publish this work in ESD, there are several issues that need to be addressed.

Thank you. We have addressed your issues individually below, and marked in the revised manuscript where we have made changes, except for updated tables and figures, or if paragraphs have been moved into a new section.

1, It is difficult to catch the highlights of this work. Many calculations have been done in this work, but by reading the manuscript, it is very easy to get lost. I would suggest the authors to make a better discussion, and the conclusion should be improved.

In light of comments by the other review we have added some extra paragraphs in the discussion/summary sections for each analysis. Given the length of the sections addressing temperature for the station plots, we have reordered these paragraphs, adding in an extra discussion on the temperatures, including references possible causes for the changes observed, before moving on to the other two variables. The summary of Section 4 now also includes aspects of the observational data and how these could have affected the results. Hopefully by including some of the figures from the supplement, as you suggest below, this also helps readers.

We have made some changes to the final summary to highlight some of the other changes further up in the manuscript, and tried to clarify this section.

2, There are many figures in the supplementary materials. But the main text discussed these figures frequently. It seems that the figures are important. Therefore, why not include these figures in the main text? Or maybe the structure of the manuscript needs to be improved. Moreover, for the figures in the supplementary materials, I would suggest the authors use “Fig. S1, S2, etc.”, to distinguish from the figures in the main text.

We were attempting to strike a balance of the number of figures in the manuscript, and not have too many to dominate the text. As we discuss up to three variables, annually and in some cases seasonally, and with a sub-daily dataset are able to split across the day as well for up to 4 moments too, we didn't want to overload the manuscript with figures.

In light of the above comment we have added a number from the supplement into the main body of the paper, but have also retained them in the supplement so that this still has a logical flow to assist readers. This increases the duplication, but we feel this is appropriate to assist readers using the supplement.

We have also updated the numbering as suggested.

3, When studying the changes, what is the statistical significance level? What method was used to do the significance test? Why use 1σ as the threshold?

In this analysis, except for the quantile regression section, we do not assess the statistical significance level of any changes. We had used the +/-1σ range to determine how reliable a trend is by whether this range includes or excludes zero, and only plot those on the maps. In light of the comments by Referee 1, and as noted in our response to their comments about Figure 3 we have reverted to showing all stations in these scatter plots. We emphasise those where the +/-1σ range excludes zero with a larger symbol as the trends are more likely to be reliable but all the stations are now plotted.

Our aim with this approach is to balance occasions where there is a large trend magnitude but also a large spread in the possible trend values from the median of pairwise slopes algorithm. There are many studies which discuss the advantages and disadvantages of any form of significance testing, and the care required to frame both the problem and the test in the correct way (e.g. Ambaum, M.H., 2010: Significance Tests in Climate Science. J. Climate, 23, 5927–5932, https://doi.org/10.1175/2010JCLI3746.1, Wilks, D.S., 2016: "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. Bull. Amer. Meteor. Soc., 97, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1, Ziliak, S. and McCloskey, D.N., 2008. The cult of statistical significance: How the standard error costs us jobs, justice, and lives. University of Michigan Press.).

We chose 1σ as this is widely used as the uncertainty of an estimated value, and use this as a way to indicate how more or less likely it is that a trend is different to zero rather than the magnitude of the trend itself. However we are actually not expecting the trends to be zero in many cases and so this has framed how we see the problem. Therefore we do not wish to add formal significance testing to this assessment.

In this study we are (a) fitting a trend to a relatively small number of points, and (b) using a linear trend to summarise changes over the 45 years of the study. We do not expect any changes to be linear, and merely use this as a way of simply quantifying changes over time, as has been done in many other studies of the past climate. Furthermore, the small number of points is the result of balancing sufficient observations per temporal bin to accurately determine the properties of a distribution with a large enough number of bins.

4, Since only data over the past 45 years were analyzed. Are the observed changes influenced by potential decadal variabilities in the climate system? Can the statistical significance test rule out the potential influences from the decadal variabilities?

As noted in the response to (3), we do not use a statistical significance test in this analysis except for the section on quantile regression. We also note that we do not expect any long-term changes to be linear, but use a linear trend to summarise these changes over time in a clear and simple way. We are in effect limited to a maximum of 46 years of data given the drop off in stations available in the HadISD prior to 1973, as discussed in Section 2. So it is quite possible that decadal variabilities could be one of the drivers for these changes in shape, and the length of data we have available may not be long enough to disentangle these effects. Our aim was to investigate what the changes in the distributions were, and point to possible causes rather than determine the most likely cause.

We have in response to comments from the reviewer 1 added discussion into the possible causes of changes in these distributions, and mention decadal variabilities there.

5, The results from this work were compared with the findings from previous studies. When the results are not in line with each other, which results are more reliable? Why? The authors may need to better explain why the results are different.

It is very difficult in this comparison to determine which study has the most reliable results. The limitations with the study done here arise mostly from the amount of data available, e.g. once sub-daily records are split into hourly anomalies and then combined in 5-year periods (Section 4). However we do not have the ability to investigate the details of other studies to pull out where differences lie - which may be from the trend fitting, the data preparation, the time periods covered, the underlying stations etc. We merely explain where our results differ from others and offer possible explanations but cannot say the exact cause or which to take.

In our re-ordering of Section 4 (from your 1st comment) we hope that we've clarified our analyses and have also added sentences outlining why we feel that we cannot say which assessment is best, especially as our opinion could come across as biased.