Reviewer #4

Comments and Suggestions for Authors

The paper analyses the long-term relationship between surface temperature and precipitation during the summer monsoon transition, including the climatology of cloud properties using MODIS data.

A year is marked as a warm or cold year if Tmax is one standard deviation above or below the 43-year mean. As there is a warming trend, there are fewer cold than warm years, which can lead to noisy statistics. Indeed the small number of even cold years means that all results are noisy, and the differences between warm and cold have further noise.

Regarding the noisy composites, and the statistical significance, please see our reply in the specific comment.

The differences in properties (especially cloud properties) between warm and cold years are not shown to be statistically significant.

Regarding the statistical significance of cloud properties, please see our reply in the specific comment.

The figure captions explain in detail how each result is computed. While this can be seen as openness, it can also mean that the captions are exceptionally long. Detailed information on diagrams is part of the analysis method and should either appear in the text or be assumed when normal practice is followed. In general much of the information in the captions is repeated in the text.

We have examined all the figure captions.
For Figures 1, 2, and 3. They are some differences in the calculations. The concept of climatology and anomaly can be assumed as normal practice, but the components used for the calculation should be stated. We have to state how we calculated the mean, anomalies, and composites, mostly because the calculation of mean, the reference state, would determine the pattern of anomalies. If we move all of them to the texts close to the figures, the reading would be interrupted. If we consider to move them to the method section, we would need to specific a calculation for a particular figure well before the figure. For example, we would need to introduce the warm and cold years for Figure 3, which need the analysis from Figures 1 and 2.
Figures 4 and 8, we believe the captions are reasonable.

Figures 10 and 11 are for the backward trajectory analysis. There are mainly three types of statistical calculations. The probability and histogram are normal practice, but we still need to state the calculation (3D) domains and time periods and simulation steps. The captions consist of these details. We did consider to move these details to the texts before our first submission. We finally placed them in the captains because we fell that they can interrupt the reading when putting in the texts.
Figures 5, 6, 7, and 9 are for those joint histograms. This is the same situation, there are some differences in the calculations of the marginal probability in different figures. Because there are lot more concerns about these figures from the reviewer, so we will not address the issues on the captions here.
Overall, unfortunately, we cannot refine all these captions in time for this revision.

Several figures have detail that can only be seen by magnifying the figure greatly.

The Figures are large enough in our uploaded version. They have shrunk in the version for review.

The English could be improved, but the meaning of text is generally very clear.

Specific comments

l8. What is meant by ‘exacerbated precipitation-temperature relationship’?
L85. Could ‘exacerbated negative precipitation-SAT relationship’ be perhaps an enhanced or exaggerated or stronger relationship?
Here we mean the intensifying responses/feedback of precipitation and temperature under the relationship. We did not use enhanced or stronger relationship because both of them may be more related to ‘close connection’but not intense responses/feedback. Normally we would say exacerbated warming or exacerbated heatwaves, therefore here I use exacerbated. However, we cannot distinguish the different between exacerbated and exaggerated.

L276. It is not clear how composites (ie averages over the whole period) can reveal information on trends during that period. In any case, both composites are noisy so that even differences between them are hard to demonstrate as being statistically significant.

We have calculated the temporal trend of Tmax and Tmin, also provided the 5% and 95% confidence intervals for a reference. We placed the figures in a newly-added Appendix A (Figure A1), and a corresponding sentence in the paragraph related to Figure 2. We decided

to place the figure in Appendix A in order to avoid the repetition of information regarding the warming trend. For this 43-year period, both Tmax and Tmin show a long-term warming trend. However, their trends are relatively weak in the first half of the period. This can be revealed as we separated the period into two sub-periods (as that for the Table 2). For the second half of the period, the warming trends of Tmax and Tmin are evident. During this latter period, 7 out of 8 cold years are found before 2012, and 6 out of 8 warm years are found after 2014. Based on their distribution over this period, the differences between the warm and cold years should be enough to indicate the warming trend. In this revision, we have modified this sentence to indicate the distribution pattern of the warm and cold years during the composite period. We wish this can avoid the confusion.

Yes, we have relatively small sample sizes for both the warm and cold years. However, given the apparent differences of the Tmax and Tmin in May between the warm and cold years, we cannot tell that all results are indeed noisy unless we have a calculation for assessing the statistical significance of the differences between the warm and cold years, and that hypothesis testing results prove that we should accept the null hypothesis of no differences between the two groups.

In this revision, we have conducted the statistical tests for the composites of Tmax, Tmin, DTR, precipitation, U850, and CF in Figure 3. The result would be upload as supplements alongside the revision. In the revised manuscript, we also added two sentences to address the result. Because this is a transition period, the temporal changes are large, so a statistical test for the whole period or the month of May could be trivial. To assess the statistical significance of the differences between the warm and cold composites, the hypothesis testing should be calculated individually for each calendar day. The null hypothesis is no difference between the warm and cold composites. In the period from 1 April to 30 June, we assess the statistical significance for each calendar day using p-value. We set three significance level (0.1, 0.05, and 0.01). When the p-value is less than one of the significance levels, we can reject the null hypothesis and accept the alternative hypothesis. The results reveal that during the May, the majority of the composite differences of Tmax, Tmin, and CF are statistically significant at level of 0.01, and that of the precipitation in the first half of May are also statistically significant at level of 0.05 in general. Indeed, the calendar days that are statistically significant are generally consistent with the days where their regions of standard errors (of sample mean) are well apart in Figure 3. This is because the standard error is one of the components used for the calculation of statistical significance. In our first submission, we plot the standard errors alongside the composite mean of each calendar day to reveal the difference between the warm and cold years based upon the implication of standard error (of sample mean). One of the reasons is that this is commonly used. The other reason is that it is relatively difficult to illustrate all the p-values for each day on the plot.

Figure 5. The colour bar in 5a does not well demonstrate the magnitude of probability; a colour bar with a tone gradient would be much more effective. In fact it is not clear that 5a is needed; 5b seems to summarise the key points well. Indeed 5c is just a different representation of the summary figure 5b. Both 5a and 5c could be removed, and so keep a focus on the key results.

Our reason to keep Figure 5a is because of the right skew (more in the smaller end) of the distribution within the moderate COT category. As for Figure 5b and 5c, although 5b can be used to show the vertical structures of different cloud regimes, the differences in different cloud types, in particular for their vertical structure changes, cannot be inferred from 5b, that is why we also show 5c. The structure changes, especially for those low cloud regimes, reasonably reflect the differences in the large-scale dynamical patterns.

Figure 6 has the same information in two different plots, and a caption that is far too long.

Our reply to this question can be found in the previous one for the Figure 5. Furthermore, Figure 6a show the differences in the vertical structures of low cloud regimes between the warm and cold years, and this is hard to see unless we display them using the stacked histograms. Figure 6c is used for comparison of each COT category.

L451. It is not at all clear that the differences in drop size distribution between warm and cold years are statistically significant.

For the statistical significance of cloud properties, our reply is not only for Figure 7, but also for Figures 5, 6, and 9. At this time, we are still not very sure how to perform a test for assessing the statistical significance for those optical cloud properties. The main reason is that these analyses are on the basis of joint histograms. We may calculate the statistical test bin-by-bin to assess the significance of the difference in the numbers of pixel. However, clouds are transient, they may just change from one bin box to another bin box. One may say that we can just test them based on the nine optical cloud types, then the numbers of low cloud detections hindered by high clouds would be a next question. As you can see, there are substantial differences in the total counts of the warm and cold years, even we have selected the same numbers of years for both. Then there is also an issue on the use of the statistical parameter for testing. For example, say if we want to assess the significance of the proportion in each bin box, then we have the differences in the total counts. In this context, even we assess the statistical significance of the differences between the two sets of total

count, one can still argue that there is a sampling issue regarding that hypothesis testing, just because of the remote sensing technical issues.
On the other hand, because currently we do not have enough knowledge to design a proper hypothesis testing to assess the statistical significance, which means that we also cannot quantitatively prove the differences in those cloud properties between the warm and cold years are statistically insignificant.
However, the distribution patterns found in the joint histograms are suitable and sufficient for revealing the structure of cloud layers and types, as well as the features of cloud mixture. Which, to a certain degree, are meaningful in association with different large-scale dynamical configurations. Our results still have the merit of revealing the cloud changes associated with the monsoon transition, as well as under a warming trend.

Again with figure 7 it is not clear that any differences in time in the relationship between COT and CER are statistically significant.

Regarding the statistical significance of cloud properties, please see our previous reply.

In Figure 8, where does the overall mean iso-height lie for each level (Figure 3)? Once again, we are unsure of the overall variability of these heights.

We have updated Figure 8, and we plotted the mean (climatology) in gray.

L528. Figure 3a does not show CTP.

This is a typo, and we have corrected it.

Figure 9. Are the differences significant? In fact the unexpected changes on the BoB suggests the changes are in the noise.

Regarding the statistical significance of cloud properties, please see our previous reply.
The changes on the BoB are actually expected. For the BoB, the development of deep convection is more active in ICP warm years, and late SCSSM onset years (in literatures). We have added the analysis for low tropospheric stability parameters in this revision. We analyzed the convective available potential energy (CAPE) obtained from CFSR reanalysis. For ICP and SCS, CAPE are larger in warm years. However, the pattern is reversed in BoB. We see this as reasonable results. Near surface hot and humid condition would result in high CAPE, the strong WPSH dynamically suppresses the development of deep convection. Some previous studies have revealed that during active disturbed periods in the tropical

monsoon region, CAPE decrease due to near-surface cooling. In short, the CAPE analysis is supportive to our results. The newly-added figure would be placed after Figure 9.

In table 2, if the bold numbers are statistically significant at some level, then this should be stated in the caption.

We have calculated the statistical significances for the differences between the correlation coefficients between 1979-2000 and 2000-2021, now the bold numbers indicate the differences are statistically significant at 95% level. As for the statistical significances for the correlation coefficient values, we stated them in the caption.

L565. The reduction in correlation between ENSO and Tmax and Tmin is likely associated with global warming, as the warming continues independently of internal climate factors.

It could be, but we cannot prove this based on the analyses in this study.

Figure 10 e and f. Are the distributions of the initial elevation of each set of back trajectories? The relevant time or period needs to be stated (rather than much of the present detail that could be assumed or found in the text).

No, they are not. The initial elevation of all the back trajectories used for this manuscript are the same (initially distributed within a layer of 1000m to 2000m). The information can be found in the method section. In addition, the periods for the collection of the back trajectories are stated in the caption. The corresponding backward time for each line is show in the legend, they are -24h and -120h. We have provided the information in our first submission.

L604. What is meant by “the specific humidity carried by the backward trajectories”? Is it the humidity at the start or end or average?

No, they are the specific humidity alongside the backward trajectories at every timestep. Which means, at every model timestep, for each trajectory, we interpolated the values of specific humidity at the nearby grids onto the location of the trajectory. These data are output every 6 h.

L607. What is meant by “Both the specific humidity and temperature exhibit statistical patterns of decreasing with heights”? They are both likely to decrease with height.

This is because there are some days of inversion pattern found in the simulation period. We did not exclude these samples therefore here we describe the statistical distribution pattern of specific humidity and temperature decrease with height. Considering that those days with inversion are really a small proportion of the total. In this revision, we modified the sentence as “Specific humidity and temperature generally decrease with height, therefore we…….”

L609. How is the temperature of a trajectory computed?

The same as the specific humidity. At every model timestep, for each trajectory, we interpolated the values of temperature at the nearby grids onto the location of the trajectory. These data are output every 6 h. Which means, there are several output parameters for each trajectory, including the location, temperature, and specific humidity. We have provided the information in the method section of our first submission.