We thank the reviewers for their valuable and constructive comments, and we thank the editor for their careful consideration of our manuscript. The feedback has greatly helped us to improve the manuscript. We have addressed the suggested changes throughout the text and modified the figures accordingly. The revisions primarily include further explanation of the CCF analysis, the seasonality of the QBO–TTL cloud relationship, the interpretation of the sensitivities, and the comparison with relevant previous studies. All changes are highlighted in the revised manuscript (using tracked changes - all revisions are shown in-line and deleted lines have strikethrough). Responses to the individual comments are provided below, with references to the corresponding line numbers in the highlighted manuscript.

R1 Comments

Major

Seasonality: This study does not consider seasonality. It would be helpful to quantify the seasonal dependency of the QBO-high cloud relationship, as shown in Sweedy et al. 2023.

Thank you for the comment. We have now looked at the seasonality, and for CALIPSO, this reproduces the results from Sweeney et al 2023. Similar to their figure 4, the following result confirms the QBO-cloud relationship peaks during MAM in the Indian ocean region, while the peak is during DJF in the western Pacific (WP).

As expected MMM is much weaker, almost 10 times weaker than observations. And the Indian Ocean peak in MAM in observation is shifted to JJA in models.

Line 329-331 in the highlighted manuscript, Text S2 and Fig. S2 in the supplementary (same as Fig. R1 below) now address seasonality.

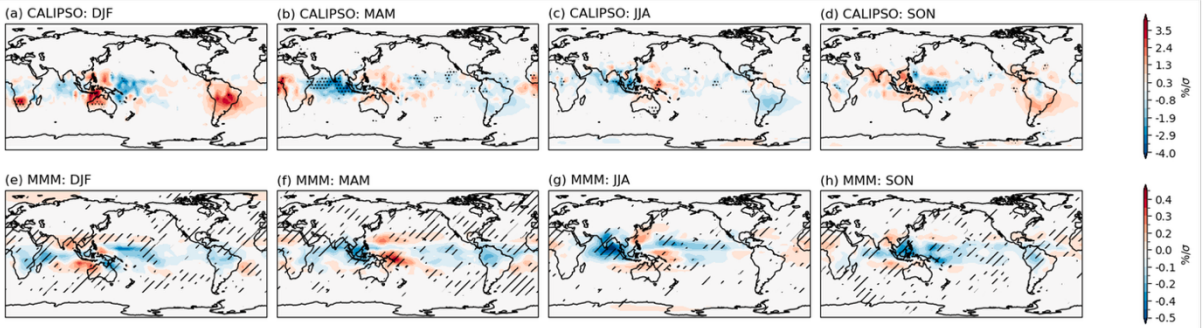

Figure R1: QBO regressed TTL cloud fraction from CALIPSO (a-d) and MMM (e-h) for DJF, MAM, JJA and SON respectively. Stippling in (a-d) denotes 90% confidence, and hatching in (e-h) denotes six out of 7 models having the same sign.

Minor

L28 Reference: You can include Haynes et al. 2021 (https://doi.org/10.2151/jmsj.2021-040)

Thank you for suggesting. We have included the reference now (see line 28).

L137 TUT: Is TUT computed in the same way as RHUT, vertically averaged within the 200 hPa below the tropopause?
Yes, that is right. We have included TUT calculation details in the manuscript now. See line 137.

L147 Eq. 1: What is r? What is the data resolution? Does 21*11 indicate latitude grids within 30S-30N and pressure levels?
Thank you for the comment. Several lines are added to CCF analysis section to clarify the details. r represents an individual grid box. The data resolution is 5°x5° as noted in line 81. 21x11 indicates the number of grid boxes in longitude and latitude. This corresponds to 105° longitude x 55° latitude. See lines 147-159.

L152 Ridge regression: I am not familiar with this method. How does it minimize the correlation between variables. TUT, S150, and RHUT are closely related with each other.
While ridge regression does not minimise the correlation, it takes care of issues with ordinary least squares regression which usually gives non-robust coefficients when the predictors are correlated. Specifically, ridge regression adds an L2 regularization penalty to the regression cost function. This penalty discourages large coefficient magnitudes and therefore reduces coefficient instability and overfitting, particularly when predictors are correlated. This results in more robust regression coefficients even though predictors are correlated. See lines 163-169.

While the partial-derivative formulation suggests that the regression coefficients of each CCF represent individual contributions (holding everything else fixed), we acknowledge that owing to correlations among the predictors, their influences may not be perfectly disentangled. Nevertheless, the comparison between, for example, Tut and Tsurface (Fig. S10) suggests that Tut has a distinct contribution to high clouds that cannot be explained by its correlation with other predictors, indicating that Tut may influence high clouds through mechanisms distinct from those associated with other controlling factors – for example as ice-cloud microphysics and cloud-top radiative cooling. Also, the temporal correlations between each pair of TUT, S150, and RHUT anomalies are mostly less than 0.25 in the tropics (e.g., see Fig. R5).

L169 Zonal wind: A 3-month lag is applied but it still shows a maximum signal at 50 hPa. Is this correct?
Yes, that is correct. The magnitude of signal at 50 hPa has reduced, but the peak is still at 50hPa. (The resolution in the stratosphere is 10hPa, 20hPa, 30hPa, 50hPa, 70hPa, 100hPa). Given that the QBO descends at a rate of approximately 2–4 hPa per month (Pascoe et al., 2005), this is physically consistent.

L245 Reference: You can include Son et al. 2017.
Thank you, we have added the reference. See line 303.

L367 Model error: Models show weaker or opposite TTL cloud sensitivities to S150 and RUHT. Why? I know this is a very difficult question. Any speculation would be helpful.

Thank you for the comments. As suggested, the exact reasons for the weaker TTL cloud sensitivities in the models remain unclear. One possibility is biases in the model mean state. Differences in the climatological temperature, humidity, static stability, and tropopause structure in the models compared to observation could lead to different sensitivities. Another recently proposed explanation is that climate models may underestimate the stability-iris mechanism (Wilson-Kemsley et al., 2025), whereby increased static stability reduces high-cloudiness through reduced convective outflow and horizontal divergence. Underestimation of this mechanism could contribute to weaker or even opposite cloud sensitivities in the models. See line 450-452.

L393 Model error: TUT signals are not consistent across models. Could this be related with model vertical resolution, lowermost level of the QBO signal in the model, or something else? Any speculation would be helpful.
Thank you for the comment. While the TUT responses differ from observations in all models (Fig. 5b), the QBO modulation of TUT is broadly similar across the models, with QBO westerlies generally associated with warming. This suggests that the model discrepancies primarily arise from differences in cloud sensitivity to TUT. For example, MRI, which more closely resembles observations, exhibits negative cloud sensitivities to TUT, whereas the other models generally show positive sensitivities. There is also some inter-model variation in the QBO modulation of TUT, which contributes to the overall spread. The latter could be associated with lowermost level of QBO signal.

R2 Comments

1. This paper would benefit greatly from more discussion/analysis related to the statistical significance. Given that we only have 14 years of calipso data, and we are trying to test something related to the QBO, we only have ~6 independent samples (not accounting for complications due to the QBO disruptions). I know this partly motivates ridge regression as opposed to linear regression, but I still wonder if you can really fit 6 predictors to the model.
   a. When showing significance, are you accounting for autocorrelation?
   Thank you for pointing this out. Currently, we haven't accounted for autocorrelation. To account for temporal autocorrelation in the timeseries, the effective sample size (N_{eff}) is estimated following Eq. 31 from Bretherton et al. (1999). N_{eff} is then used in the calculation of the t-statistic and the corresponding p-value is obtained from the t-distribution with N_{eff}-2 degrees of freedom. As an independent check, the significance of the regression coefficients was also evaluated using a heteroskedasticity- and autocorrelation-consistent (HAC) estimator (Newey and West, 1987). The HAC-based results showed broadly similar spatial patterns of statistical significance. In the figures showing linear regression with the QBO index, statistical significance is indicated by p-values calculated using N_{eff} to account for temporal autocorrelation. See lines 203-209.

   Regarding fitting 6 predictors to the model, we are calculating sensitivities of clouds to the selected CCFs, and hence we are learning relationship from all monthly variability and not just from the QBO.

We have compared the CALIPSO results with observations with slightly longer record (Figures R2 and R3). While CALIPSO only has 6 QBO cycles, MODIS and ISCCP have longer time ranges with 22 and 34 years respectively. These observational datasets also give similar results as CALIPSO for QBO regression on TTL cloud amount (Fig. R2), and the CCF contributions (Fig. R3). We note that ISCCP keeps all unclassified cloud retrievals in 0.3-1.3 optical depth bin in the top altitude. Thus, this bin is removed from the calculation.

It is also reassuring that, in most models, regression analyses performed on 20-year segments show consistent major signals across segments (see the unhatched areas in Fig. S9 in the updated supplementary, where most segments agree in sign), suggesting that short 20-year segments are able to capture the major signals.

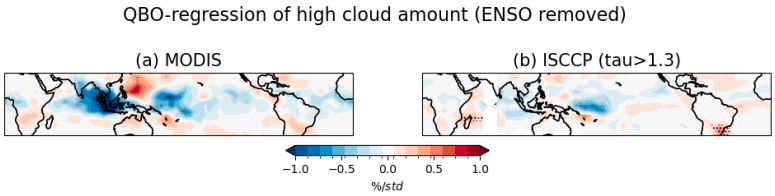

Figure R2: QBO regressed TTL-cloud fraction (<440hPa) from (a) MODIS and (b) ISCCP

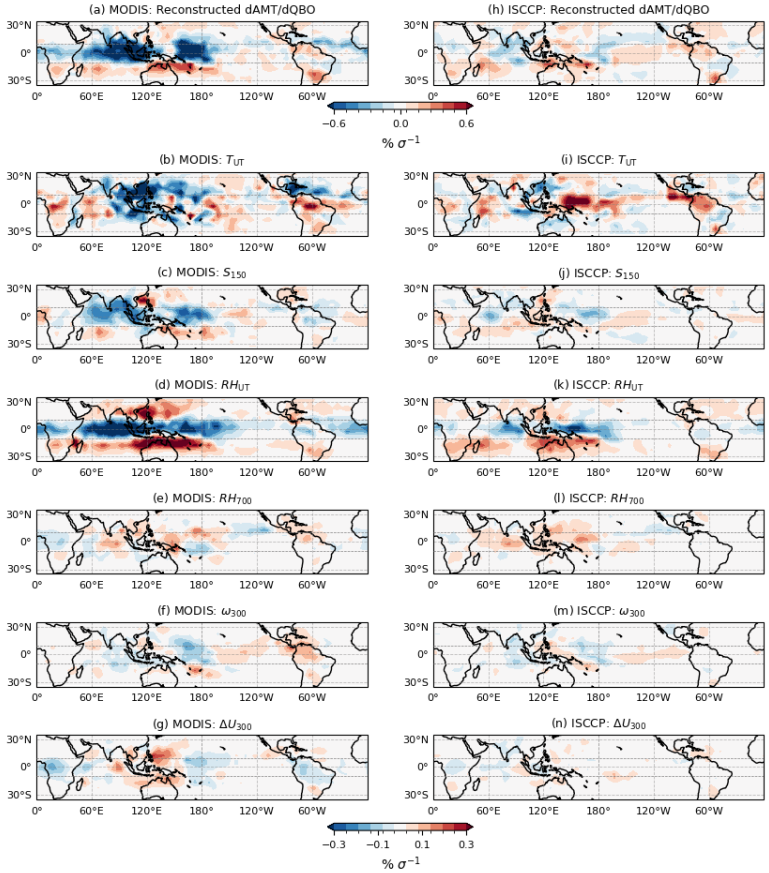

Figure R3: Reconstruction of QBO-regressed TTL cloud fraction anomalies using CCF analysis for (a) MODIS and (h) ISCCP. (b-g) Contribution of each CCF to the MODIS QBO–TTL cloud relationship and (i-n) to the ISCCP QBO–TTL cloud relationship.

2. Line 124: How is ENSO regressed out? Is it regressed out of the QBO index and all the latxlon datasets? Did you regress it out with a lag in the ENSO index?

ENSO impacts on TTL clouds have been found to maximize after a few months (Tseng and Fu, 2017).

ENSO is regressed out of both QBO and all other latxlon datasets by regressing out the nino3.4 index without any lag. We have included text in Data and Methods to describe how ENSO is regressed out. While the ENSO–cloud relationship may involve a small lead/lag (Fig. R4), removing ENSO with zero lag was sufficient to highlight the QBO signal, which is the focus of this study. Moreover, since multiple variables from the troposphere and stratosphere are considered, applying a single lead/lag to all variables would not be appropriate. See line 196-201.

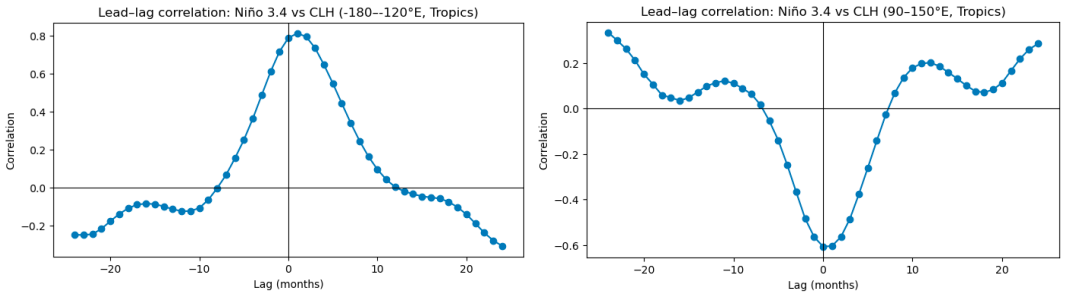

Figure R4: Lead–lag correlation between the Niño 3.4 index and regionally averaged high cloud fraction over (a) 180°–120°W and (b) 90°–150°W.

3. How is TUT defined? Consider more clearly showing how you calculate TUT, especially given its relevance in the later results.

   Thank you for pointing this out. We have missed to include the definition. We have added it to data and methods section now. See line 137.

4. Line 139: You change upper tropospheric stability to S150 related to the Chen and Thompson results. I am wondering if you should also change w300 and u300 to things more relevant for the TTL. For example, you might change w300 to w100 or u300 the shear between 300 and 100 hPa. Given that many of the thermodynamic changes driven by the QBO operate via changes in the vertical velocity in the upper-troposphere and lower-stratosphere, the w100 might be interesting to look at.

   Thank you for the interesting suggestion. We have repeated the analysis with w100 and in observations, the reconstruction slightly improved. There was little impact on the tropical mean reconstructed value of d(AMT)/d(QBO), which went from -0.47 to -0.48. While the contribution of w increased (+0.07), this was offset by decreases in the contributions from Tut (-0.019) , RHut (-0.013) and S150 (-0.024).

   We note that alternative CCF sets may provide an improved reconstruction of the QBO-cloud relationship. However, exploring other CCF sets is outside the scope of this study and the reconstruction using current CCF set performs well with a spatial correlation of 0.87 to QBO regressed TTL cloud amount (Fig. 3a compared to Fig. 2c). See lines 344-347.

5. Equation 1: I found it difficult to understand what each term/symbol represented. Consider defining everything in the sentence immediately following. For example: "… where AMT represents the cloud fraction, r represents the individual grid cell, Θi represents the sensitivity of high cloud fraction to the ith cloud-controlling factor (Xi) and M is the total number of cloud-controlling factors (M=6)".
   Thank you for the feedback. We have added the suggested clarifications now. See line 147-159.

   1. Writing the equation for Θi as d(high_cloud)/dXi would have been helpful for me.
      Thank you for the suggestion. We have added this to equation 1.

6. Fig. 1: I liked the comparison of models and observations here. It may be helpful to provide a difference row as well as the obs and modeled results.
   a. Consider reducing the hatching thickness as the current version feels too crowded
   b. Consider adding thin black lines as another contour at fixed levels to highlight changes in the magnitude
   c. For Fig S2, consider using the same colormap and colorbar as done in Fig 1 for ease of comparison
   d. For Fig S3, consider using same colormap and colorbar as Fig 1

   Fig. 1 now has thinner hatching. We have also reduced the number of levels in the colormap, so that contours are more distinguishable. Fig S2 and S3 (moved to S5 and S6 in the updated supplementary) are replotted with the same colormap as Fig 1. The difference plot for Fig. 1 is added in the supplementary (Fig. S7).

7. L224: I noticed this at this point, but this is true throughout the paper. Consider reducing the emphasis on QBO westerlies reducing high clouds. While this is of course true, some studies have shown that the impact of the QBO on the TTL is actually much stronger during QBO easterlies (Tegtmeier et al., 2020). This is more subtle in this study because it is regression based, but may confuse the results. For example in L227-229, you state that the zonal asymmetries in the cloud impact are tied to QBO westerlies, when in reality, these are just the regression coefficients and may be dominated by easterly QBO variability
   Thank you for pointing this out. We have now added a description in the Data and Methods section line 192-195. We have also tried to reduce the emphasis on QBO westerlies throughout the paper by replacing with 'a westerly QBO anomaly'.

8. L294-295: This is a very interesting statement. I would like to understand this better. If this suggests that clouds respond separately to these 3 predictors, could you give more intuition or examples on how? More importantly, how do we square these results with the recent results of Chen and Thompson (2026) which seemed to argue that previous studies overstate the QBO-cloud coupling, and that studies should only look at the S150 variable? Is the flaw in that study that

they only consider S150? If so, it would be helpful to have some physical intuition on why TUT is important. I know you expand on this slightly in L300-305, but having more discussion would highlight that stability alone is insufficient to describe cloud changes, and highlight the discrepancy with the Chen and Thompson study.

Thank you for the comment. We have now added a discussion in the 'conclusion and discussion' section to how our results compare with Chen and Thompson study. To summarise, we show that other mechanisms beyond the stability pathway matter for the QBO-cloud relationship, such as temperature influencing ice cloud microphysics or cloud top radiative cooling. Tseng and Fu (2017), for example, mention ice particle sediments and cooling associated with wave propagation as plausible mechanisms for temperature-related TTL cloud changes. However, further investigation is needed to better understand the underlying mechanisms. See lines 531-540.

1. I think Fig S6b shows that including all predictors with TUT (Fig. 5b) gives a larger response than all predictors with TSURF (Fig. S6b). Please correct me if I am wrong.

   That is correct. This confirms that T_{UT} provides a contribution independent of S_{150}.

2. How much of the variance between S150 is shared with Tut?

   Thank you for the comment. We are not entirely sure what the reviewer is referring to, but we assume that they are asking about the correlation between S150 and Tut. Figure R5 shows the spatial map of temporal correlation between S150 and Tut. This correlation isn't particularly strong in most of the tropics, suggesting that the two variables have some independent variability (and therefore potentially some independent impact on cloud).

   The correlation flips sign across the tropics, probably due to where Tut lies relative to S150, since S150 is defined at a fixed pressure level, whereas Tut is an average from the tropopause to 200 hPa below it. From Fig. R6, the tropopause over the Indian Ocean is generally at a lower altitude than over the rest of the tropics. As a result, the TUT calculation includes a larger number of pressure levels below 150 hPa, which may contribute to the negative correlation between S150 and TUT in this region. Interestingly, even at locations where S150 and TUT are negatively correlated, both variables exhibit a positive correlation with the QBO index (not shown).

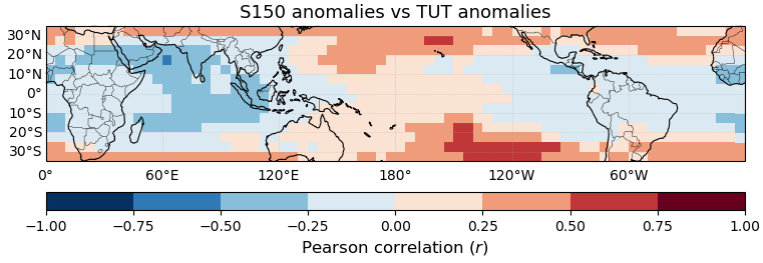

Figure R5: Spatial distribution of the correlation between TUT and S150 anomalies after removing the ENSO signal. The removal of ENSO results in only negligible changes in the correlation strength

9. Fig 4: Given the importance of TUT here, it should be included in this figure.
Thank you for the suggestion. While Tut is an important term, the simplified sensitivity map can be misleading because it is it a reduced-dimensionality form of the original 4D sensitivities (one longitude-latitude map per target grid box). We have added a description in the main text now to clarify this, and instead of keeping the simplified Tut sensitivity map, a new figure S1 showing the regional sensitivity map for at a sample grid box has been added. See lines 436-445 and Text S1.

    1. Fig S7: I don't understand the TUT sensitivity. Is it true that the reconstruction in Fig. 3b should be obtained by multiplying the pattern in Fig. S7aa? If I understand correctly, the positive values in Fig. S7a would suggest that an increase in temperature leads to an increase in cloud fraction. This seems opposite of what I would expect, and also contradictory to the results of Fig. 1. Please let me know if I am reading this incorrectly.
    Fig. 3b is related to the sensitivity shown in Fig. S7a and d(CCF)/d(QBO) in Fig. S7aa. However, these are not directly multiplied at each grid point. For each grid point, the sensitivity is defined over an associated 21 × 11 grid-box domain, as described in the methodology (also see response to previous comment). This sensitivity field is multiplied by the corresponding 21 × 11 d(CCF)/d(QBO) field and summed over the domain to obtain the reconstructed value at that grid point. The sensitivities shown in Fig. S7a are the spatially averaged sensitivities over this 21 × 11 domain. An example of sample grid box sensitivity profile and details of calculation (Text S1 and Fig. S1), which confirms that there will be reduction in cloud fraction as seen from Fig 3b, is included in the supplementary file now. See line 464-470. Also, please note that Fig. S7 is moved to Fig. S11 in the updated supplementary file.

    2. Also, the TUT sensitivity in S7 seems spatially disorganized compared to S150. I think more interpretation of the results here would be helpful
    Thank you for the comment. The text added to the CCF analysis results, text S1 and Fig. S1 now address this question. As can be seen from Fig. S1a, Tut signals are negative in the deep tropics and positive outside, and averaging this across regional domains results in a weaker, patchier sensitivity map. See Text S1 and line 436-445 in the main text.

10. L345-355: Could some of the zonal discrepancies between obs and models here be explained by differences in tropopause between obs and models? In observations, the QBO influence is much stronger above the tropopause, and much weaker below. If the tropopause is incorrect in models, this might prohibit

QBO anomalies from extending into the troposphere. In observations the tropopause has longitudinal dependence, and this may be true for models as well.

Thank you so much for the suggestion. We have looked at the tropopause pressure level zonal profile (Fig. R6), and models consistently have a slightly lower tropopause pressure level compared to observations, although the zonal structure seems to be reasonably captured. It is possible that the difference in tropopause height leads to small differences in the vertical structure of the QBO influence on stability. See line 419-421.

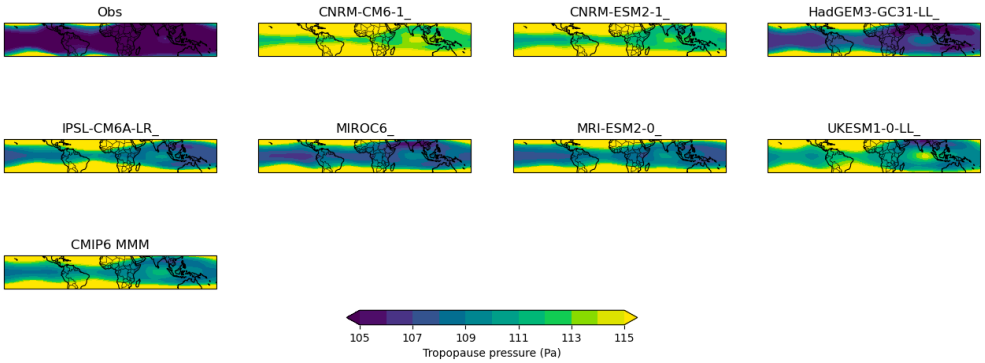

Figure R6: Latitude-longitude profile of climatological tropopause pressure in ERA5, the 7 models analysed in the study and the CMIP6 MMM.

11. Fig 5: Should the observation total in Fig. 5b be the same as the observational value in 5a? I am not familiar with the regression techniques used here, but if not, then does this change the interpretation of the role of TUT from Fig S6?
Thank you for the comment. Ideally, the total responses in Figs. 5a and 5b would be identical. However, some differences are expected due to limitations in the reconstruction, such as the possible contribution of missing CCFs, or nonlinearities in the relationships. The aim is therefore to maximise the similarity between the reconstructed and observed responses. When Tsurface is used instead of TUT, as in Fig. S6 (Fig.S10 in the updated file), the reconstructed tropical-mean response is only −0.38, whereas including TUT improves it to −0.47, compared with the observed value of −0.54 in Fig. 5a.

12. L370-380: I noticed this here, but I think it is relevant throughout the paper. I am having trouble interpreting the %/std(QBO) units here. Sweeney et al. (2023) show variations in cloud fraction of about 2-5%. I know this is a composite difference compared to a regression coefficient here, but it still seems that there might be a discrepancy. Please correct me if I am misunderstanding.
Thank you for highlighting this. We have added a brief description in the main text now. See lines 186-190.

The difference mainly comes from using seasonal versus annual means, and representing results in units of per standard deviation of the QBO. Sweeney et al. (2023) use a −8.5 to 8.5% colour-bar range in their Fig. 4 for longitudinally and seasonally resolved QBO responses. A similar range is obtained for the observed CALIPSO data used in this study when calculating the QBO westerly-minus-

easterly composite difference as shown in following Fig. R7.

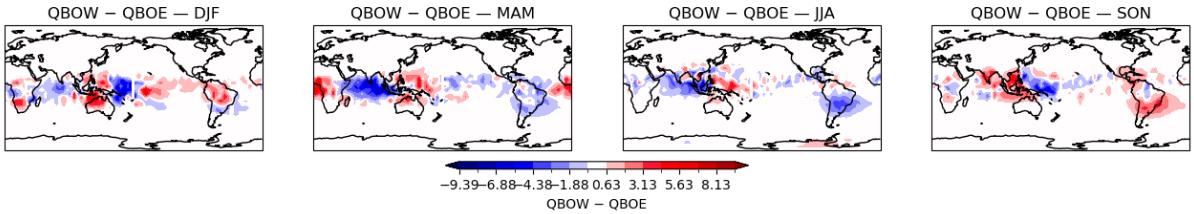

Figure R7: Seasonal composite of TTL cloud fraction anomalies, calculated as differences between QBO westerly and easterly phases (westerly minus easterly).

However, taking the seasonal mean reduces this colour-bar range by approximately half, as positive and negative anomalies shift spatially between seasons. Furthermore, the mean QBOW and QBOE winds are approximately 5.8 and −11.5 m /s, respectively. Thus, a composite difference represents the cloud-fraction change associated with an approximately 18 m/s change in QBO wind. In contrast, our regression shows the cloud response per one standard deviation of the QBO, approximately 10 m/s, further reducing the colour-bar range to approximately −2.5 to 2.5%, as shown in Fig. R8. In the main text, a narrower colour-bar range is used to better highlight the MMM signals.

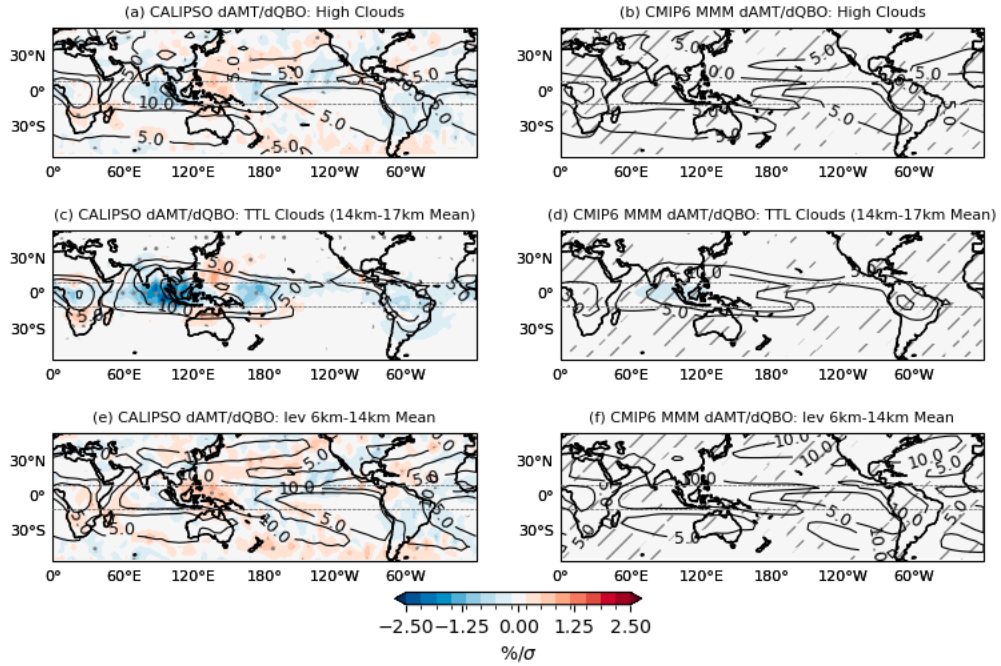

Figure R8: QBO regressed high-cloud fraction (<440 hPa), TTL cloud fraction (<180 hPa) and 440–180 hPa cloud fraction in CALIPSO (a) and CMIP6 MMM (b), represented in units of percent cloud fraction per unit standard deviation of the QBO index. Overlaid line contours indicate the respective climatology cloud fraction. Stippling in (a,c,e) denotes 90% confidence, and hatching in (b,d,f) denotes six out of 7 models having the same sign.

Minor Comments:

L87: "...and so on" → "among other reasons (cite relevant paper)"
Thank you. We have edited. Chepfer et al., 2010 cited on the next line addresses the details. See line 86.

L98: "…underrespresentation of very thin clouds in CALIPSO-GOCCP compared to standard CALIOP-NASA products" → This seems like it could be important given that many of the TTL clouds are very thin. Is this something that may impact the results?
Compared with Sweeney et al. (2023), which uses the standard CALIOP–NASA product to calculate QBO composites, QBO-composited cloud fraction values from CALIPSO-GOCCP are comparable (see Fig. R7 above), suggesting that the choice of CALIPSO product does not substantially affect the results.

L159: Discussion on the QBO definition might be better for section 2.
Thank you, we have moved it to section 2.2.

L231: consider citing Tegtmeier et al. (2020)
Thank you for the suggestion. While Tegtmeier et al. (2020) looks at QBO amplitude in temperature and CTP, in L231 we were focusing on the spatial profile of QBO-regressed high cloud fraction as shown in Fig. 2a, which is similar to Fig. 6b in Chen and Thompson (2026) or Fig. 9 in Gray et al. (2018). Apologies if you were referring to another paper or a different line.

L255: "with a proposed mechanism…" → "with one of many proposed mechanisms involving…"
Edited. See line 313.

L257: "Figure 2 indicates the strongest QBO-cloud relationship in observations occurs over the maritime continent…"
Edited. See line 315.

L257: Consider also citing Lin and Emanuel (2023)
Added. See line 315.

Fig 3: Given that RHUT is one of the three main contributors, consider moving it to row 3.
We have moved it to the 3rd row showing CCF contributions in Fig. 3.

L311: "…do not align spatially in the MMM…" Do you mean that they do not align spatially with the observations or that the 3 predictors have less similar spatial patterns than those of the observations?
In observations, contributions from all three predictors are in the same region. In models, the contributions from these predictors have different spatial profiles.

L322-323: My understanding is that this has been known from previous papers, consider citing them here.
We are not sure if there is any literature that investigate both UTLS and surface variables to conclusively establish the QBO-TTL cloud relationship mechanism is in-situ; we are unable to find any.

We have made all the requested edits below.

L91–92: "(Swales et al., 2018**). has been developed" should read "...2018) has been developed."
L92–93: "generates outputs that is comparable" → "that are comparable."
L155: "20 years of MODIS data are also used from comparison" → "for comparison."
L257: "the QBO–cloud relationship in observations peak occur over maritime continent" → "...peaks over the Maritime Continent"
L355: "Figure S6 (first raw)" → "first row."
L401–402: "both cloud sensitivities and QBO modulation play part" → "play a part."
L403: "the cloud sensitivities to TUT and S150 varies" → "vary."
L104: "22 years of MODIS dataset are analysed" → "22 years of MODIS data are analysed." Also consider avoiding starting the sentence with a number.
L111: "global circulation models (GCM)" I am familiar with GCM meaning general circulation models, not global.
L24, L122 and throughout: inconsistent unit spacing ("10hPa," "150hPa," "440hPa") vs the spaced form used elsewhere ("150 hPa"). Standardize.
L53: "1 W/m2" → "1 W m^{-2}" (and superscript).
L124: "Niño3.4" → "Niño 3.4."

R3 Comments

1. Introduction/conclusion: The authors cite studies such as Sweeney et al. (2023) and Chen & Thompson (2026) which are closely related, but could more clearly state what the original contribution that their research represents. What are the main points of difference? What is the significance of the emphasis on TUT in the authors' results?
   Thank you for the comment. We have added two paragraphs in the conclusion to discuss the results with respect to previous studies. Our observational results agree well with the findings of Sweeney et al. (2023), while we further implement CCF analysis to investigate the meteorological pathways linking the QBO to TTL clouds and to assess CMIP6 model biases. Our CCF analysis shows that the contribution from S150 is consistent with the S150-related contribution identified by Chen and Thompson. However, whereas Chen and Thompson argued that the remaining signal in the total regression may not be physical, our results indicate that a larger fraction of the total regression signal can be explained when additional relevant meteorological variables, such as TUT, are included in the analysis. This suggests that additional pathways may exist through which the QBO modulates TTL clouds, such as direct thermodynamic effects, including influences on ice-cloud microphysics (Tseng and Fu, 2017). Further investigation is required to identify such mechanisms.

   See lines 523-540. Also see text S2 and Fig. S2 for seasonality comparison with Sweeney et al. (2023).

2. Robustness testing and details of CCF analysis (section 2.1):
   1. The authors give a useful explanation of the replacement of surface temperature with TUT and the use of 150 hPa static stability instead of a

mean upper-troposphere value. However, these variables are not defined in sufficient detail, especially TUT which is central to the conclusions.
Thank you for the comment. We have added the definition now. See line 137.

2. The authors claim that the changes to the CCFs "better represents" QBO-related pathways (L138) but without justification. Can you quantify the reconstruction skill, for example the fraction of variance explained or a spatial map of residuals? Can the authors show these metrics for a few different sets of CCFs to convince the reader that the chosen set really is the best choice?
We have now included a supplementary figure (Fig. S3; same as Fig. R8) showing the spatial pattern of d(AMT)/d(QBO) and reconstructions using three sets of CCFs. One with original CCF set used in the study, one with stability changed to S_mean and one with temperature changed to Tsurface, as used in Wilson Kemsley et al., 2025. The spatial correlations between QBO regressed TTL cloud amount and three sets of CCF choices are analysed. And the relative amplitudes of the tropical mean reconstruction (%) are also calculated (Fig. R9).

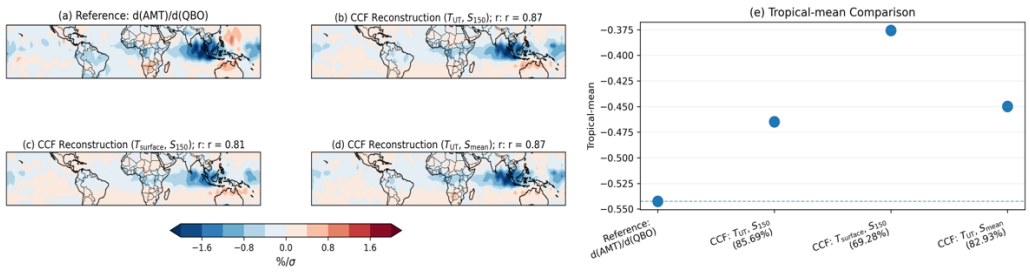

Figure R9: (a; same as Fig. 2c) QBO-regressed TTL cloud fraction and (b; same as Fig. 3a) its reconstruction using the CCF set described in the study, (c) the CCF set used in the study but with TUT replaced by Tsurface, and (d) the CCF set used in the study but with S150 replaced by Smean. (e) Tropical-mean QBO-regressed TTL cloud fraction and its reconstructions. The values in parentheses next to the reconstruction labels indicate the percentage bias in the tropical mean.

Using $T_{surface}$ instead of T_{UT} reduces spatial correlations and tropical mean relative amplitude (%) considerably, while S_mean vs S_150 difference mainly comes in deep tropics.

However, we acknowledge that alternative CCF sets may provide an improved representation of the QBO–high-cloud relationship and identifying the optimal CCF set warrants further investigation. Building on previous studies, the current selection provides a good reconstruction of the observed spatial pattern.

See lines 342-347.

3. Methodological details.

1. The methods section is already reasonably long (in that it takes a long time to get to what are relatively succinct and neatly presented results) but there remains a lack of detail in the methodology. For example, many of the terms in the CCF definition (equation 1) are not defined, and it is difficult to interpret given the choice of variable names and symbols. The authors might consider comparing-and-contrasting their methodology with other closely-related studies such as Chen & Thompson (2026) – the main point of differences appears to be the use of area-averages rather than grid-point values; can the authors elaborate on why?
   Thank you for the comment. Based on this suggestion and comments from other reviewers, we have modified the description of the CCF analysis. Ceppi et al. (2021) showed that although the results are not highly sensitive to the exact choice of domain size, restricting the sensitivities to local grid-point relationships substantially degrades the CCF reconstruction. Figure S1c shows the sensitivity of cloud fraction to S150 at a sample grid box, illustrating the influence of surrounding grid boxes. See lines 147-159.

   Despite differences in the calculation of sensitivities, the spatial pattern of the stability contribution to the total d(AMT)/d(QBO) reported by Chen and Thompson (2026) is similar to that obtained in our analysis.

2. Similar to above, there is no information given on the ridge regression methodology or the way in which ENSO is removed. Did the authors consider removing other important modes of variability? The authors may consider adding an Appendix with the fine details of the methodology included.
   Thank you for highlighting this. For the ridge regression analysis, we follow the methodology of Ceppi and Nowack (2021), as described in their Materials and Methods section on cloud-radiative sensitivities. We have now added a few sentences clarifying how ridge regression differs from multiple linear regression and directed readers to Ceppi and Nowack (2021) for further details. See line 163-169.

   We have also added a short section (section 2.2) to the Data and Methods describing the QBO and ENSO indices and how ENSO is removed. See line 179-201.

   While ENSO is a prominent mode of tropical variability with a characteristic timescale close to that of the QBO, other modes, such as the PDO, AMO, and MJO, have characteristic timescales substantially longer or shorter than that of the QBO and are not removed since they are less likely to interfere with the QBO signal.

3. The authors use a standard QBO index (L162) at 50 hPa with a three-month lag. Since the QBO/high-cloud pathway operates in the upper-troposphere/lower-stratosphere, have the authors considered using a

QBO index at a lower pressure level with no lag? Are their results robust to changing the QBO index?

Thank you for the comment. QBO index at 50 hPa was selected based on previous studies and considering that in some of the models QBO wind signals barely come down to 70hPa. For comparison, QBO-regressed TTL cloud anomalies in observations using the 70 hPa QBO index with zero lag are shown below (Fig. R10). While the regression amplitudes are nearly twice as large, the spatial and temporal patterns remain unchanged. While a 70 hPa index would be preferable physically, considering different QBO depths in models and as the pattern remains unchanged, the 50hPa QBO index is used. The details are now included in the main text. Line 181-184.

The following figure showing lead-lag correlations also confirms that the peak correlation between QBO index at 50 hPa and high clouds is when 50 hPa winds are at a 3-month lead, while it is at 0-lag for the 70 hPa wind index (Fig. R11).

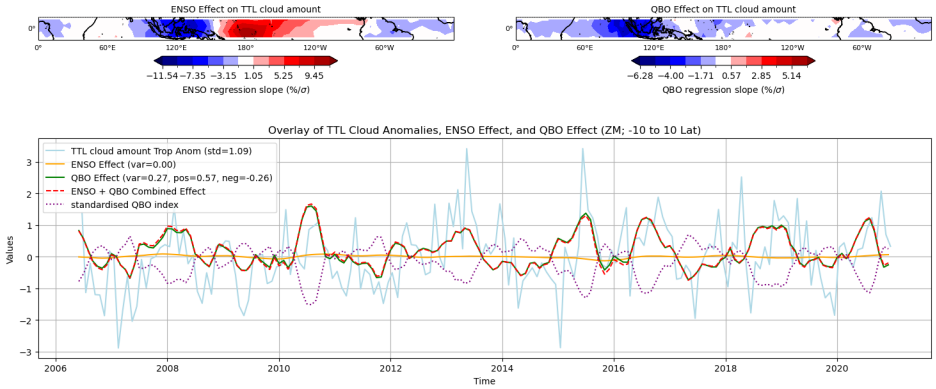

Figure R10: Tropical TTL cloud amount anomalies over time. ENSO and QBO contributions are shown as yellow and green lines respectively. TheQBO index used is zonal-mean zonal wind at 70 hPa averaged between 10S and 10N.

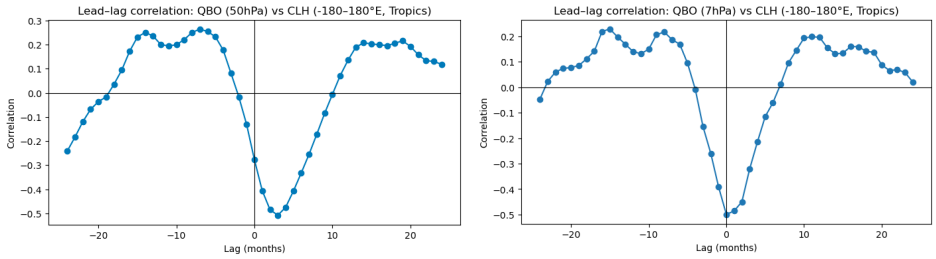

Figure R11: Lead-lag correlation between QBO index at 50 hPa and 70hPa respectively with high cloud fraction.

4. The CALIPSO record used is short, containing only a few QBO cycles, and importantly the record includes QBO disruptions. This is an important caveat that should be more clearly dealt with by the authors.

Thank you for pointing this out. To address this, we have compared the CALIPSO results with observations with a slightly longer record (Fig. R2).

While CALIPSO only has 6 QBO cycles, MODIS and ISCCP have longer time ranges with 22 and 34 years respectively. These observational datasets also give similar results as CALIPSO for the TTL cloud amount regression on the QBO. We note that ISCCP keeps all the unidentified cloud fraction in the 0.3-1.3 optical depth bin in the top altitude. Thus this bin is removed from the calculation.

It is also reassuring that, in most models, regression analyses performed on 20-year segments show consistent major signals across most segments (see the unhatched areas in Fig. S9 (in the updated supplementary), where most segments agree in sign), suggesting that short 20-year segments are able to capture the major signals.

5. The authors state (L294-295) that clouds respond separately to TUT, RHUT, and S150. I am not sure I follow their arguments. Can the authors clarify this statement?
   Thank you for the comment. The coefficients obtained from ridge (multiple) regression can be interpreted as partial derivatives, meaning that they describe the impact of each controlling factor individually, while holding all other factors fixed. However, given that the predictors can be correlated, we acknowledge that their influences may not be perfectly disentangled. Nevertheless, the comparison between, for example, Tut and Tsurface (Fig. S10 in the updated supplementary) suggests that Tut has a distinct contribution to high clouds that cannot be explained by its correlation with other predictors, indicating that Tut may influence high clouds through mechanisms distinct from those associated with other controlling factors – for example as ice-cloud microphysics and cloud-top radiative cooling.

6. The authors state (L112) that 7 models generate a “reasonable QBO”. What is the selection criteria used?
   For models with CALIPSO data available, a fast Fourier transform (FFT) is applied to the 50 hPa zonal-mean zonal wind (Fig. R12 right column). Models exhibiting a clear QBO signal are then selected. A sentence is added to clarify this. Line 111-112.

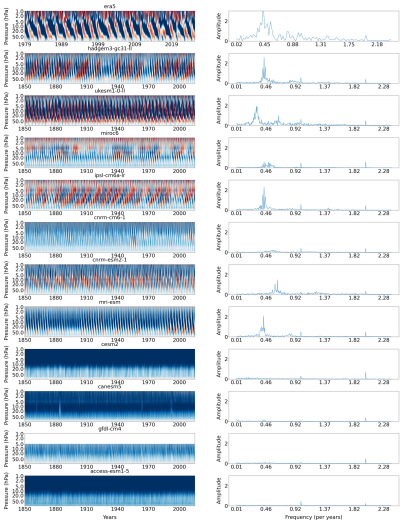

Figure R12: Time evolution of the zonal-mean zonal wind from ERA5 and CMIP6 models, together with the corresponding Fourier spectra at 50 hPa, showing the dominant periodicities of the QBO.

7. Presumably the multi-model mean is an even-weighted mean of the 7 models (L113). Is that appropriate? Are the models all independent in the sense that they use distinct dynamical cores? If not, the authors might consider weighting the average.
   Yes, the study uses an equally weighted mean across the seven models. Although some models share the same dynamical core, the QBO–cloud relationships can differ substantially between them, as illustrated by HadGEM3-GC31-LL and UKESM1-0-LL (Fig. S9 in the updated supplementary). While the models used here are not fully independent, determining suitable weights goes beyond the scope of this study, and unweighted multi-model means are a standard choice in the literature.

Minor comments

1. L53: you mention the seasonality of the QBO impact on high clouds in the introduction, but do not explore this in your own study. Is this a simple extension of the results that the authors might consider including? Particularly since they mention consequences for the QBO-MJO connection (L255) which is strongest in boreal winter.
   Thank you for the comment. We have now calculated the seasonality and this is added as a supplementary plot (Fig. S2). The observational results are similar to results from Sweeney et al. (2023), with a DJF peak in the west Pacific and MAM peak in the Indian Ocean. While models on average capture the DJF signal correctly with a west Pacific maximum, the Indian Ocean maximum is in JJA instead of MAM as seen in observations. And the regression coefficients in MMM are almost 10 times weaker compared to observations. See line 329-331 and text S2.

2. L97-98: “results in an underrepresentation of very thin clouds in CALIPSO-GOCCP compared to standard CALIOP-NASA products” -> what is the consequence of this in the context of this study?
   Compared with Sweeney et al. (2023), which uses the standard CALIOP–NASA product to calculate QBO composites, QBO-regressed cloud fraction values from CALIPSO-GOCCP are comparable, suggesting that the choice of CALIPSO product does not substantially affect the results.

3. L332, 432: I found use of the term “mislocated” confusing. It is hard to discern whether the authors mean that the cloud sensitivity to the CCFs is wrong itself, or whether the erroneous sensitivities result in erroneous responses to the CCFs that produce a spatially distinct response in the CMIP6 MMM compared to observations.
   Thank you for raising this. We have modified the term to ‘biases in spatial distribution of cloud sensitivities’.

4. Figure improvements
   1. Fig. 1: the stippling and hatching makes it difficult to see the underlying plots. Either reduce the stippling/hatching or use a colourbar with more contrast. Can the authors provide some guidance to the reader on how to interpret the units (e.g. m/s/sigma)?
      Thank you for the comment. We have now reduced the number of levels in the colormap, so that contours are more distinguishable, and reduced hatching and stippling. A description of units is now included in the Data and Methods section. See lines 186-190.

   2. Fig. 3: should share the same colourbar as fig 2 to aid comparison. There is also a *lot* to absorb from this figure: could it be simplified to help with interpretation somehow? Note that a lot of the results shown in this figure are not even mentioned in the text, so arguably not relevant to the central story.
      Thank you for the comment. The total reconstruction in Fig. 3a, h, and o uses the same colour bar as Fig. 2. A narrower colour-bar range is used for the individual CCF contributions to better highlight the MMM signals. Following another reviewer’s suggestion, the RHUT panel has also been moved above so that the major contributions are grouped together. We hope these changes improve the clarity and representation of the results.

5. Typos:

   The following typos are now corrected.

   1. L91: “mimic” -> “mimics” and remove the . after the citation
   2. L156: “from comparison” -> “for comparison”

3. L195: “The reduction … in CALIPSO (Fig. 1i)” -> Fig. 1e?
4. Fig. 2 caption refers to MODIS but figure subtitles say CALIPSO
5. Fig. 2 caption says hatching in “(b, d, e)” -> (b, d, f)?
6. L257: “peak occur” -> “peaks”
7. L297: “with S6). And” -> join to form one sentence?
8. L355: “first raw” -> “first row”
9. L454: “CMIP5/6 data” -> was CMIP5 data used?
10. L416-417: “but systematically underestimating” -> “underestimate”