This paper presents a comparison of the total column water vapor from the ESA TCWV-COMBI data set to ERA5 and a suite of CMIP6 models. TCWV is analyzed in terms of PDFs' percentiles and composites on large-scale circulation. The study shows that applying a consistent cloud screening is important for comparing the different data sets. Over oceans, both ERA5 and CMIP6 models agree well with the observations, regarding variability and relation to circulation. Over land, larger differences and discrepancies occur, likely related to changes in instruments.

I think this is an interesting comparison paper of a new observational data set with new model simulations and reanalysis, and suitable for publication in ACP. I have no major concerns, only a few minor and specific comments below, which hopefully will help to improve the paper.

Minor comments:

1. English language: I had the impression that the language and wording could still be improved. I've just listed a few typo/wording issues below which I came across, but would recommend a careful final check.

2. Cloud screening: I understand from the description in L110ff that the cloud screening of the model data (for both CMIP6 and ERA5) has been adjusted to optimize the comparison to TCWV-COMBI, with the additional constraint to have enough data included - and that this screening is critical for the comparison. I find this description somewhat qualitative and insufficient. I'd recommend to describe the screening approach in a clearer way (perhaps even including a figure for illustration), to enable readers to repeat the analysis.
Specific comments:

L114ff: How sensitive are the results with respect to these screening thresholds?

Table 2: Why Eyring et al. (2016) as reference for ERA5? Wouldn't Hersbach et al. (2020) be more appropriate?

Figure 2: Why is there this change in variability in TCWV-COMBI in 2010? There is much less variance in the later period.

Figure 2: Are the outliers in Tropical-land TCWV-COMBI realistic (especially in 2004)? I doubt that TCWV can almost double from one time to the other. I recommend to include a cautionary note.

L186: I'd recommend to present the final sentence of this subsection ("The bimodal distributions ..."), explaining the cause of the existence of bimodality, earlier in this section (e.g., in L176).

Figure 3: It could be mentioned already regarding Fig. 3 that the observations show higher frequency of moist extremes.

Figure 4: I'd prefer some color scheme centered around 1 such that 1 is represented by white, similar to the one used in Fig. 7. This would make the anomalies clearer. (Similar for Figs. 5/8/10).

L222ff and Figure 6: I found the description of Fig. 6 here somewhat oversimplified. E.g. over trop. ocean the TCWV-COMBI data set shows moistest regimes not for upward motion, but for weak subsidence. I was wondering about the reasons, and also think this should at least be mentioned and discussed in the text. Also, CMIP6 and ERA5 seem to agree somehow better in terms of TCWV(w500), than they compare to observations - any ideas for possible reasons?

L254 and Figure 8: Interestingly, all data sets except ERA5 show a dry anomaly around 2008. Any idea why ERA5 disagrees in that respect? Perhaps the formulation in L254 should be more careful, saying that the 2008 dry anomalies for ERA5 are only observed for subsidence and weak upward motion regimes.
L292: I think this should read "moistening trend" here, if it refers e.g. to Fig. 7.

L312: I'd add: ... and the GCMs over oceans.

L317: Maybe better: "This clearly shows the necessity/importance of having access to the simulated water vapour ..."

Technical corrections:

L11: absent in?
L16: a critical role
L64: investigateS
L64: One "the" too much.
L65: participated in?
L67: Blank after "datasets".
L82: is discussed
L95: participating in?
L127: considerable --> considerably
L129: ERA5
L130: Comma after "Hence".
L136: approaches --> approaches
L137: evaluateS
L161: Blank after "TCWV".
L163: Better "time series" instead of "climatology"?
L163: Fig2a --> Fig. 2a (Please check throughout the paper that references to figures are according to ACP style.
Fig. 2, caption: cover
L193: reddish
L233: Blank after "atmosphere".
L238: Blank after "different".
L240: Blank after "models".
L247: regime
L249: interval
L259: Blank after "variabilities".
L291: tipping
L291: associated with
L296: associated with
L298: ...ESA TCWV-COMBI data and ERA5...
L303: Blank after "factors".
L307: differences
L314: I think a parameter can not be "easier".