"The roles of the Quasi-biennial Oscillation and El Nino for entry stratospheric water vapour in observations and coupled chemistry-ocean CCMI and CMIP6 models" by Shlomi Ziskin Ziv et al. submitted to ACP

1. Summary

Ziv et al investigate the roles of the QBO and ENSO for the interannual variability in entry stratospheric water vapour in observations and chemistry climate models outputs from CCMI-1 and CMIP6 using their Multiple linear regression and different supervise learning regressions (named here Machine learning). They compare the ability of their own MLR with the different supervise learning regressions to evaluate their MLR robustness compared to their different supervise learning regressions in capturing the interplay between the QBO and ENSO influences to water vapor entry.

The results idea is of great interest to potential readers and worth it for publication. However, the manuscript has 3 mains issues, which are a lack of honest motivation of the study, methodological failure and finally the presentation of the result issue that I have detailed in the major and specific comments. Major revisions are needed in order to make the paper suitable for publication. There are some additional points that need to be clarified. I apologize if I misunderstood something.

2. Major comments

i. *Honest motivation of the study*: The concern is the main motivation of the paper. From Page 2, L13-35 and Page 3, L1-10, the discussion is not clearly and honestly reported. The authors are using the result from Garfinkel et al. 2018 based on CCMs (geosccm) to argue about questionable "nonlinear" ENSO response. The result of "nonlinear" ENSO impact on water vapor is still an open question as the observations does not show this nonlinearity. In addition, the models used in Garfinkel et al 2018, 2021 are based on spontaneous generated QBO, which does not reach the tropopause and does not have the same QBO phases and strength as the observations. This lead to wrong water vapor modulations by the interplay between the QBO and ENSO in the lower stratosphere, therefore, to questionable result of Garfinkel et al 2018, 2021. The discussion about about "El Nino and La Nina can lead to moistening" is still very questionable, knowing the inability of CCMs to reproduce a descent tape recorder (Keeble et al 2020). The look like circling discussion (e.g. dog biting its tail) about the ''nonlinearity'' of ENSO need to be discussed clearly as it still needs to be proven with the observations rather than it's already an established results. As we know CCMs have several issues when it comes to water vapor entry in the lower stratosphere. Finally, one important remark is the Authors are using their OWN multiple regression model, which is not the same as the Dessler et al. 2014, Diallo et al, 2018 and Tian et al. 2019. There are as many as different MLR in terms of predictors, including a dynamical or fixed lag and solver. Just note that the Diallo et al. 2018 method is not a simple MLR where you can predefine a fixed lag as your regression but a multivariate hybrid method. In order word you have used your OWN regression, therefore, you should be that general as even your regression has issues.

ii. *Methodological failure:* The regression model used here is failing to reproduce the ENSO (El Nino and La Noina as well) induced impact on water vapor variability (Figure 6). The QBO coefficient also looks strange. The ENSO-square even looks like a second QBO coeficient. Actually, the ENSO impact on H2O structure is a horseshoe pattern as shown in Konopka et al 2016 and Avery et al. 2017. Apparently, the regression model used here only is not capturing ENSO look-like impact on H2O entry. When multiplying QBO impact by

ENSO impact (QBOxENSO), the result looks like an ENSO impact on H2O pattern. The Figure 7 and the large residual over the entire period (trained and tested) in Figure 9 both corroborate the failure of Ziv et al MLR model. The QBO signal from their MLR is also questioning. So, major analysis are need to investigate this failure before concluding. Possible diagnostics are: First, it would be great to see how well your MLR and ML are able to capture the altitude-time cross section of the tropical H2O variability induced by the QBO and ENSO (5S-5N mean of their effect). Second, estimate the R-square error of the residual. Third, verification of the used ENSO proxy if it is not too small and also especially in the manuscript (Page 7, L2), you stated using NINO3.4 from ERSSTv5 data with a 1981-2010, while the analysis period is 1994-2019. Regarding the ML, the different supervise learning method are barely described in the manuscript. The other main issue is the training and testing period of the Machine learning. The authors did not use an independent training date set for testing the performance of the machine leaning model. The approach of using the same data randomly sampled and divided into 5 fold won't help to assess the performance of the ML model. This is a serious issue. You should show at least show the ML performs in the unseen test sample to disclose over-fitting issues etc. The lag used in the manuscript is not clear if it is observed one or the one from the CCMs. Please clearly describe each method and explain what you have done. Finally, the cold point temperatures are very well negatively correlated with the H2O as the latter is determined by freeze-drying process (Fueglistaler & Haynes, 2005; Fueglistaler et al. 2013; Poshyvailo et al 2016; Grandville & Birner al 2016). The CPT as H2O then are both modulated by the climate modes of natural variability, including QBO and ENSO. So comparing CPT and the QBO and ENSO as predictors is not making sense at all. Since early findings, we know the strict relationship between H2O and CPT. One should use the CPT if one would like to predict H2O entry but when it comes to separating and understanding different contributions to H2O inter-annual variability, it does not make sense.

iii. *Presentation of the result issue*: The structure and presentation of the results have issues which need to be improved. The authors discussed about CCM2 (Page 4, L14-18), while they are not using it. I recommend to remove this part but clarify the CCM1 representation of QBO (nudged or spontaneous) and SSTs (modelled or observed), which missing here. For instance, EMAC has also the nudged QBO, which is not mentioned, but you emphasise the WACCM water vapor coefficient are due to the nudged. So it should be the same for EMAC bot no. In addition, the level of 82.54hPa used here is not a reference level, knowing that model like WACCM has a high tropopause (about 90 hPa). I would recommend to do these analysis of the manuscript at one fixed level 70 hpa for all data sets, which is actually the reference level where tropospheric influence is separated from the stratospheric ones. They could interpolate all the data at the 70hpa level.

3. Specific comments

a) Page 2, L9, Please add citations: Punge et al. 2009, Niwano et al. 2003 & Diallo et al. 2018.
b) Page 2, L13-20, please discuss the zonal mean struture of the ENSO induced impact on H2O based on the observation that has been found in previous litterature (Randel et al. 2009, Calvo et al 2010, konopka et al 2016). This is what so far the truth.
c) Page 2, L21-30 please rephrase the entire paragraph. The claimed ``nonlinear ENSO impact on H2O" still need to be proved in the observations, therefore, it should not be presented as ground true the same models are pointed out having issues with the QBO, which stuck at 50hPa, not realistic QBO phases compare to observed one. Conclusions from these that struggle to reproduce the tape record should be take with caution, which is not the case here.
d) Page 2, L34-35 please remove the citations "Diallo et al. 218; and Tian et al. 2019" as they are not simpl MLR as you frame here.

e) Page 3, L3, this statement “First, Garfinkel et al 2018 found ...ENSO is nonlinear” needs to rephrase and made clear by precising it is model based and not consistence with the observations finding yet.
f) Page 3, L1-10, please discuss also these papers: Evans et al 2014; Brinkop et al 2016, Less et al 2012, Diallo et al. 2018 about the interplay between the ENSO and QBO impact on H2O entry.
g) Page 4, L1-10 please precise that you are using the CCMI phase 1 models. In addition, please explain the model issues about getting the QBO right in the CCMI-1 and CMIP6 models.
h) Page 4, L14-27 please remove the CCM-2 discussion. It is confusing the reader as any way you are focussing on CCMI-1. Please emphasize the models ability in reprodcuing ENSO and QBO impact on the tape recoder and the uncertainty that induces in the H2O entry.
i) Page 5, L1: Please do the analysis at 70hPha for all the plots.
j) Page 5, table 2, please the QBO and SST infor mation for each model in the table.
k) Page 6, the captions of Figure 1 are not very clear. Please clarify them.
l) Page 6, the Figure 1 should be done at 70hPa for all models and observation.
m) Page 7, L1-2, please clarify "...ERSSTv5 data with a 1981-2010 base period".
n) Page 8, L5-15, a clear description of the different supervised learning regression are need here to improve clarity of the method.
o) Page 9, L3-4, please remove the citations Dessler et al 2014 and Diallo et al. 2018 as you are not using their models or out put of their models for comparison. In addition, your regression model has issues in reproducing the ENSO and potentially QBO impact structure on H2O (Figure 6); tape recorder plot of QBO and ENSO induced impact on H2O and has large residual too.
p) Page 10, L1-8, the approach used here to test the performance of the model is an issue as it you're not test the ML on unseen data for test set. How the overfitting or under fitting issues are evaluated then? It would be great to add a figure in the main paper or supplement about the ML performance showing trained period and unseen predicted H2O period. Please clarify also the training period.
q) Page 11, L13-24, Here the authors should not generalise about the MLR and its results but precise it is THEIR MLR with its limitations. The whole paragraph nee to be revise after evaluating the ability of their MLR to capture QBO and ENSO induced impact on H2O as altitude-time tropical cross-section.
r) Page 11, L25-34, the SHAP method comes out off blue. Please clarify and rephrase the paragraph
s) Page 13, the coefficient of their MLR in Figure 6a & b are wrong as well as the Figure 7. ENSO impact on H2O is not similar to classical method results. Please evaluate clearly, why? In addition, the Figure 6 d e.g. ENSO squarred is very likely a QBO signal as you are not using two QBO index with a chosen lag for all latitude bin this may impact you MLR results. The MLR needs to be evaluated before drawing any useful conclusion here.
t) Page 13, L1-8, QBO being predominate in modulating H2O entry have been already found by Diallo et al. 2018 and confirmed by Tian et al 2019. Please discuss them.
u) Page 14, L2-5, knowing the model inability of reproducing the QBO down to the low stratosphere, it is a bit strange that the author aims at evaluating the model ability to capture the interplay between QBO and ENSO impact on H2O entry. Please rephrase the entences.
v) Page 17, L16, the zonal structure temperature and H2O anomalies find in previous studies (Randel et al 2009, konopka et al 2016) is a result of the averaged between a region of updraft (cold) and subsidence (warm).