

Atmos. Chem. Phys. Discuss., referee comment RC1  
<https://doi.org/10.5194/acp-2021-972-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Review

Anonymous Referee #1

---

Referee comment on "The roles of the Quasi-Biennial Oscillation and El Niño for entry stratospheric water vapor in observations and coupled chemistry–ocean CCM1 and CMIP6 models" by Shlomi Ziskin Ziv et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2021-972-RC1>, 2022

---

The paper focusses on the factors affecting the interannual variability of stratospheric water vapor entry in the tropics in observations, CCM1 and CMIP6 models. The authors contrast the use of a variety of techniques: multiple linear regression and 3 machine learning methods. Cold point temperatures are the main factor explaining the water vapor variability. They discuss the merits of the different techniques and the relative importance of the QBO and ENSO. They also find non-linear interactions to be important. The comprehensive models, whilst will suffering from a QBO that is not deep enough, have nonetheless improved. The paper is well written and provides an good description of machine learning techniques applied to a geophysical problem. The figures are also mostly clear.

### Specific comments

(1) Make it clear earlier during the introduction that you are looking at interannual variability and not the seasonal cycle.

(2) Some of the CCM1 model have multiple ensembles. Do you average over all of them? If so, does this result in less variability and thus make it harder to compare to those runs with only 1 ensemble?

(3) In the figures, would it be possible to have the models with a nudged QBO labelled in bold text? It would make identifying them easier.

(4) On line 4, page 6, you mean ERA5/ERA5.1 I think?

(5) On page 6, line 11, "Note that the correlation of the BDC with the QBO is -0.66 (Figure 2), and hence including both in a single regression or ML model can lead to overfitting." I disagree with this statement. Multicollinearity in your predictors causes a variety of problems but does not specifically cause overfitting. See page 283, Applied linear statistical models 5th edition by Neter et al. (2004). Your validation stage should show if overfitting is an issue.

(6) Page 10, line 15, the non-linear predictors are interesting but I struggle to relate them to physical processes. Could you give the reader a sense of what ENSO<sup>2</sup> might be?

(7) The values in Figure 6 are somewhat hard to read. Could you add a few labelled contour lines please?

(8) Figure 7 feels unnecessary since the same information can be conveyed with the text.

(9) In figure 9 (a to c), the text suggests that the solid black lines are observations (and they are not described in the caption) but where are there two parts and at different values? Label the models in 9(a).

### **Minor comments**

Page 1, line 164, Emissions

Page2, line 5, through the its

Figure 1. Labels are a bit small and hard to read.

Figure 4. Are the units of the H2O anomalies correct?

Figure 5 and Figure 9. You use "std" and "std dev". Choose one to be consistent and also explain the abbreviation in the caption.

Figure 5(a) I am confused about the histogram. Is it normalised? If so, why are the values  $>1$ ?