We thank Referee #3 for the very helpful comments and suggestions. Here are our replies with which we hope to clarify some misunderstandings.

- General reply: Reading the comments of the three referees, we have the impression that we probably raised wrong expectations with respect to what we achieved with our study. All three referees agree in revealing weaknesses in our argumentation, which we believe can be explained and therefore eliminated:

  - Part of the problem is probably caused by misunderstandings (maybe due to unclear formulations), as for instance the precise definition of "drop" (sudden decline versus period of low values thereafter), or the presumable influence of "nudging" as we applied it.
  - Another, more severe part, is in fact caused by the lack of important information, which we erroneously hold back (mainly to shorten the manuscript), although it was found by our analyses. Most importantly, we did not put sufficient emphasis on our simulation RC1SDNT (i.e., nudged, but without mean temperature nudging), which was only mentioned briefly, but not discussed in full detail.

  We hope that we can clarify our findings with additional information and revision of the text, where the misunderstandings occur. The details about that are outlined in our point-by-point replies below.

- *This paper explores the ability of models to capture the post-2000 drop in stratospheric water vapor, and the factors that led to the drop. The authors find that a specified dynamics version of the model can capture the drop, while a free-running model with observed SSTs and a QBO nudged to observations grossly underestimates it but can capture some elements of it. They then argue that El Niño/La Niña and the QBO were crucial forcing mechanisms for the drop.*

  Reply: This is a good summary of what we did. However, it possibly contains a first misunderstanding, namely about term "post-2000 drop". If this means the "sudden decline of water vapour in 2000", i.e. the period until its minimum is reached, the referee is right. If, however, the ≈5 year period of low water vapour is meant, it is a misunderstanding. We never claimed that the QBO can explain this 5 year period. All we say is that QBO is essential for the sudden decline.

  To clarify this, we will define (and use) two different phases of "the drop" in our revised manuscript as:

  - Phase 1 is the short period of the steep decline between the drop onset (i.e., its maximum) and its subsequent minimum. The difference between max. and min. will be called "amplitude" of the drop.
  - Phase 2 is the period of low values between the minimum and the start of the recovery.

Furthermore, we will clarify in the revised introduction which of the phases are addressed and discussed in which Section. For instance, Sections 3 and 4 are about the millennium drop (phase 1 and 2), whereas Section 5 analyses only the phase 1 of other "drops".

We use a hierarchy of 4 different model setups to analyse the millennium drop, i.e. the sudden decline in 2000. We find, that a nudged setup (RC1SD) performs best. This cannot be expected a priori for water vapour, since the hydrological cycle is freely evolving.

A nudged setup excluding the mean temperature from nudging (RC1SDNT) also reproduces the millennium drop, however, with a smaller amplitude. This is related to the cold point temperature bias.

Next, a free running simulation (RC1) forced with observed SST shows a similar onset of the drop but largely under-represents the amplitude, caused by an even larger cold point temperature bias. Note that the observed SST used here is very similar to the SST used when nudging is applied, and thus can be excluded as cause.

Last but not least, a free running simulation with simulated SST (RC2) shows no drop at all. This is a result that does not surprise, because the dynamical situation is not related to the observed (or reanalysed).

The analysed gradual degradation of the drop signal from RC1SD and RC1SDNT over RC1 to RC2 is further enhanced (or manifested) by the difference in the QBO signal between the different simulations (see Figure 14 for RC1SD and RC1). Note that the QBO at roughly 90 hPa is key for the temperature signal affecting water vapour, i.e., at an altitude where the QBO nudging strength is already reduced and therefore relies on signal propagation.

The nudging procedure is solely used to reproduce the observed (or reanalysed) synoptic scale situation (i.e. meteorological patterns), which cannot be reproduced by a free running setup, even if forced with observed SSTs. The water vapour, however, is in all cases developing freely. Further, nudging does not correct for model errors as long as the global mean temperature nudging is not included (see above). Thus, with a hierarchy of simulation setups (from free running (forced with simulated SSTs), forced by observed SSTs, nudged w/o T-mean and nudged with T-mean) we are able to analyse the influence of different drivers. To our knowledge, this has not been done with other GCMs or CCMs so far. Our new finding is, that the drop itself is only in parts masked by a model error, namely the cold point temperature bias.

- *I found this work to be somewhat unconvincing. If SSTs were so important, then both the free-running model and the specified dynamics version should show the millennium drop. While the lower stratospheric QBO is weaker in the free-running version as compared to the specific dynamics version, and thus the model is under-representing this pathway, it is difficult to draw* *conclusions as to the importance of the QBO unless additional simulations are performed in which the QBO does propagate far enough downward. Finally, the weak drop in the free-running simulation doesn't last as long as the drop in the specified dynamics simulation, and part of why the millennium drop was so interesting is its >5 year duration.*

Reply: This might be another misunderstanding due to our formulations. We still think that the SSTs are important, because the simulation RC2 with simulated SSTs (i.e., not related to real SSTs) does not show any signature of the drop (Figure 4). In contrast, the free running simulation RC1 forced by observed (or reanalysed) SSTs does show some characteristics of the drop (Figure 4): phase 1 is partly represented, e.g., the timing of the onset (i.e., the maximum water vapour right before the fast decline) is almost correct. However, the drop amplitude is underrepresented (i.e., the minimum is too large) compared to the nudged simulations (RC1SD and RC1SDNT). Also phase 2 is visible, but the duration is indeed shorter and the minimum is too large.

Our conclusion is therefore that the correct SSTs are important to trigger the drop (i.e., phase 1) and also, at least partly, for the period of low values in phase 2. The absolute value of the water vapour anomaly minimum during phase 2, however, cannot be explained solely by SSTs.

We agree that "it is difficult to draw conclusions as to the importance of the QBO", if this is meant in a quantitative sense (e.g. as the QBO contribution to the drop amplitude) from the simulated millennium drop period only. All we find and discuss, however, is that common to all phases 1 of all analysed drops in other years, is the fact that the QBO is coincidently changing from west to east phase. As shown in Figure 14, this correlation is weaker in RC1 compared to RC1SD, because the QBO timing in RC1 is different from the "real" timing. This occurs, despite the applied QBO nudging for two reasons: First, the nudging does not force a one-by-one representation of the nudged data by the model, the applied relaxation time was 58 days. The model thus still develops its own dynamical state. Second, the relevant altitude is in the "nudging transition" region, meaning that the direct effect of the nudging is even weaker and the QBO signal depends more on the signal propagation from above.

Nevertheless, we see a clear QBO oscillation in all simulations and in the observations of temperature and moisture. Therefore, qualitatively it is doubtless that the QBO modulates water vapour in all simulations. The question we cannot answer, however, is how large this effect is. We see that during some periods the QBO temperature effect does not propagate as far down as during others. As we show for the phase 2 of the millennium drop, the QBO signal is partly compensated by an increased upwelling which causes a lower cold point temperature. In RC1 this effect is weaker compared to RC1SD. To illustrate this, we will add a Figure to the new supplement (see also below), which shows the time evolution of zonal mean temperature and water vapour anomalies versus height for simulation RC1SD, i.e., a similar figure for RC1SD as Figure 8 for RC1.

In the revised manuscript we will clarify our discussion and conclusions accordingly.

- *General Comments on Content: A. Fundamentally, it is unclear to me how the QBO and ENSO could even potentially be the answer to the millennium drop, as both of them have a characteristic timescale (2.5 years and ≈5 years respectively for a full period) that is shorter than the duration of the drop (>5 years). Any given ENSO event lasts one or two years at most, and stratospheric memory for a quantity like water vapor is on the order of months, so it isn't clear how ENSO could even mechanistically lead to a long-lived drop. Stated another way, any drop that lasts longer than 5 years must be driven by a process that can persist in a given phase for 5 years. It is worth noting that there is not a single long-lived (>5 year) drop in either RC1 or RC2 in figure 4. IN addition, the composite analysis in section 5 also suggests that the events are of relatively short duration (at most two years). (That being said, the millennium drop in figure 4 in RC1 does seem to last for 4 years, so there is some hope. There are modes of SST variability that last longer than ENSO.)*

  Reply: This is in our opinion a misunderstanding due to different meanings of "drop". We argue with ENSO and QBO only for the above defined phase 1 of the drop and before, **not** for phase 2! This will therefore be clarified in the revised manuscript.

- *B. My intuition based on previous work is that ENSO and the QBO are important for changes in stratospheric water vapor, and probably contributed a big chunk of the drop for least a couple of years. In terms of ENSO, two publications not cited should be discussed in the manuscript:*

  Garfinkel, C. I., M. M. Hurwitz, L. D. Oman, D. W. Waugh (2013), Contrasting Effects of Central Pacific and Eastern Pacific El Niño on Water Vapor, GRL, 40, Stratospheric4115-4120, doi: 10.1002/grl.50677 Garfinkel, C.I., D. W. Waugh, L.D. Oman, L. Wang, and M.M. Hurwitz, (2013). Tem- perature trends in the tropical upper troposphere and lower stratosphere: connections with sea surface temperatures and implications for water vapor and ozone, Journal of Geophysical Research: Atmospheres, 118(17), 9658-9672, doi: 10.1002/jgrd.50772

  The first paper demonstrated that La Niña leads to moistening of the stratosphere, while the impacts of El Niño were dependent on the specific nature of the El Niño event (some lead to dehydration, others t' little effect in the annual mean). This paper is entirely consistent with the authors' arguments, as they find that large drops follow La Niña events when the stratosphere is moistened. Note that this is somewhat in contrast with the analysis of

Dessler, A.E., M.R. Schoeberl, T. Wang, S.M. Davis, K.H. Rosenlof, and J.-P. Vernier, Variations of stratospheric water vapor over the past three decades, J. Geophys. Res., 119, doi:10.1002/2014JD021712, 2014

who find that warmer mid-tropospheric temperatures lead to more stratospheric wa- ter vapor. This point should be discussed in more detail in the revised manuscript, specifically near line 24920:15-20.

The second paper shows that SSTs have led to a dehydration trend over the historical record, and more relevantly, to a period of enhanced dehydration in the early 2000s (that is weaker than suggested from satellite/balloon products). This second paper is also consistent with the present analysis. However, both of these papers as well as the authors' RC1 simulations indicate that SSTs are not the full answer to the millennium drop.

Reply: Thank you for this comment! We will include the discussion on these publications in our revised text.

- *In terms of the QBO, the authors claim that the QBO is crucial, but don't provide the analysis to convincingly demonstrate this. The present experiment will (by design un- fortunately) miss some of the influence of the QBO. Figure 14 strongly indicates that the QBO in the lowermost stratosphere is mis-represented and much too weak in the RC1 experiment, while the QBO at these levels is likely crucial in order to capture the effect of the QBO on water vapor. I strongly suggest that the authors perform a modified RC1 experiment in which the QBO nudging is strong enough so that lower stratospheric winds mimic those observed. It would be very interesting to compare such a revised RC1 experiment to the present one to see whether the QBO does, in fact help with explaining the magnitude of the drop.*

  Reply: Here, we probably have the same misunderstanding as above. We only claim that the "timing of the QBO" is crucial for the phase 1 of the drop. We just wanted to point out, that in our RC1SD simulation a coincidence between drop phase 1 and QBO phase change (west to east) is present. This has also been reported by Dessler et al. (2014) as being important to generate large amplitudes in water vapour.

  Nevertheless, we see the gap in our reasoning for phase 1: Figure 14 shows that the QBO anomaly is more pronounced in the nudged simulation RC1SD (left) compared to the free running simulation (right). Indeed, this is not the only difference, because also the absolute cold point temperature is different, because the nudging in RC1SD includes also the nudging of the mean temperature implying a bias correction. In our revised analysis, we will include also the millennium drop of the RC1SDNT simulation, in which the temperature bias is not corrected.

  We agree, however, that additional sensitivity studies are required to corroborate our findings and mention this in our revised "Summary and Conclusions".

- *C. I found the manuscript somewhat tedious to read, somewhat repetitive, and difficult to follow. I have several suggestions for how to improve the text below, but I suggest that the authors carefully edit the paper before submitting their revised version.*

  Reply: We will carefully edit the paper!

- *D. On a relatively minor note, the bottom row of figure 3 doesn't appear to be consistent with figure 1. Figure 1 suggests that the RC1SD integration is quite good at capturing the length of the drop, but the bottom row of figure 3 gives a gloomier picture.*

  Reply: This is indeed a misunderstanding, again related to the usage of the different drop phases. We guess, in your comment you refer to phase 2 of the drop when you say "length of the drop". Figure 3, however, shows the duration of the phase 1 in unit "months". The confusion is most probably caused by the word "length". We will clarify this in the revision.

  Moreover, Figures 1 and 3 are based on differently combined data sets: In Figure 1 RC1SD is compared to the HALOE/Aura-MLS data and in Figure 3 to the HALOE/MIPAS data. Last but not least, Figure 3 shows the result of a new analysis (as explained in Appendix A4) which includes the folding of the model data with a remote sensing average kernel.

- *Minor comments: 24911:2 the first sentence of the manuscript is very unclear*

  Reply: We will reformulate it.

- *24913:5 section 5 is about ENSO and the QBO (i.e. contributors to the drop). Section 6 is a discussion.*

  Reply: We will correct this in the revised manuscript.

- *24914:21 'in water vapour we supplement the EMAC simulations with a combination of satellite observations . . .'*

  Reply: Will be reformulated.

- *24915:26 to my eye, both temperature and water vapor are captured quite well. Can this be quantified via a correlation analysis?*

  Reply: We will calculate the Pearson's correlation coefficients (between model results and observations) for cold point temperature anomaly and water vapour anomaly, respectively and add the results to the Figure.

- *24916:24 Figure 3 is introduced quite abruptly. How was this figure constructed? I think reference to the appendix is necessary (assuming I understood the appendix).*

  Reply: Yes, indeed! We will expand Section 2.2 and introduce all data sets used first. We will also reformulate the first sentence introducing Figure 3 to clarify this additional evaluation of the year-2000 drop characteristics (water vapour strength, length and drop date).

- *24917:25 this discrepancy between water vapor and temperature is very confusing. Section 5 'attributes' this to the QBO (as far as I can tell), but it is hard to believe the analysis in section 5 considering the poor quality of the QBO in the lowermost stratosphere.*

  Reply: We guess, you refer to simulation RC1 here, in which the water vapour anomaly does not follow the temperature anomaly as direct as in RC1SD. However, we do not claim in Section 5 that this is due to the QBO. We attribute this rather to the bias in cold-point temperature. At least that is what we intended to say. We will recheck Section 5 and eliminate misleading arguments pointing to the QBO for this aspect. Parts of this misunderstanding are maybe also related to the confusion with the different drop phases.

- *Figure 6: I suggest removing the RC2 curve. It doesn't contribute in any way to the authors' points.*

  Reply: It seems to be a misunderstanding, but RC2 is not presented in Figure 6. In Figure 6 we present the moisture anomalies from RC1SD, RC1 and RC1 with the eruption of Mt. Pinatubo. The red curve thus shows the effect of additional heating of the stratosphere on the water vapour variability. We removed this curve, because this was also suggested by another referee.

  In case you are, however, referring to Figure 7: We are very hesitating to remove the result for RC2, because we need the results of our hierarchy of 4 simulations, in order to disentangle some of the effects. Note that in RC2 the simulated SST is completely unrelated to the observed, however, the QBO is nudged and therefore its phase correct.

- *24924:18 'we experience' is the wrong word*

  Reply: We will change it to 'we find' in the revised manuscript.

- *24925:28-24926:25 This is somewhat long-winded and tedious. The authors' point is that the model is missing processes that are potentially important. This could be stated more concisely.*

  Reply: We agree. We will reformulate and shorten this part in the revised manuscript.

- *Section A4: I assume this is for figure 3. This should be stated explicitly*

  Reply: We will refer from the figure caption to Appendix A4 and likewise from the text in A4 to the Figure 3 in the revised manuscript.

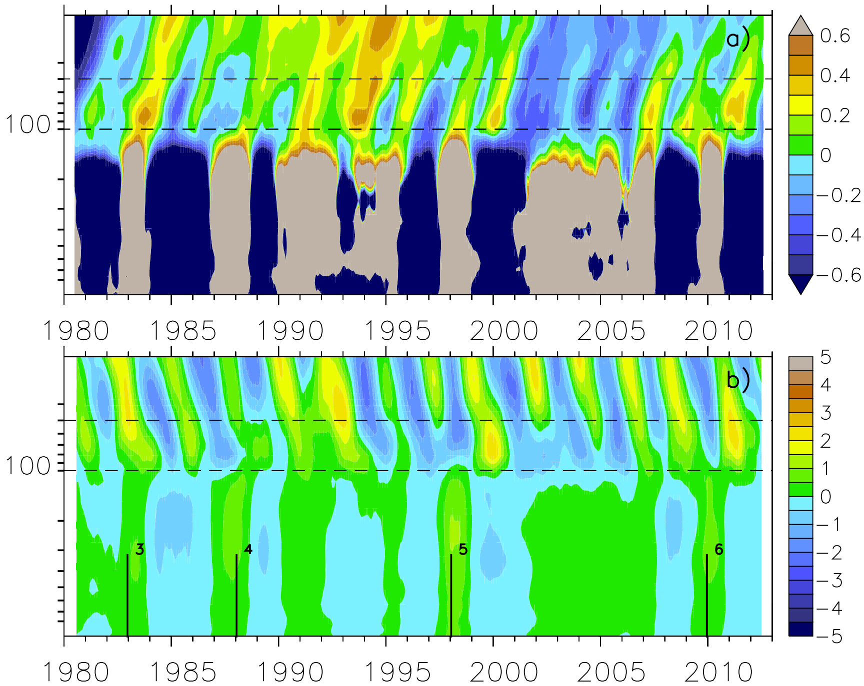

Figure 1: **(a)** Temporal evolution of moisture anomalies (ppmv). **(b)** Temporal evolution of temperature anomalies (K). RC1SD simulation, 10° S – 10° N, (12 month running mean). Strong El Niño events are labelled. The altitude range covers the pressure levels from 900 to 30 hPa. The dashed lines mark the region between 100 and 50 hPa.