We thank Stephan Fueglistaler for the very helpful comments. 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.

- *Brinkop et al. present a study of the sudden drop of water entering the stratosphere in the year 2000 using the Chemistry-Climate Model EMAC. The model is forced with observed SSTs, and the QBO is imposed by nudging with observed stratospheric winds. An additional run where the model is nudged against ERA-Interim is presented. The paper is generally well written, with some technical details requiring clearer description (outlined below). My main concern with the paper is that the numerical model results presented do not support the key statements in the text - I am looking forward to reading their rebuttal. First, Figure 4 shows clearly that only the fully nudged model run (which I must assume to be almost identically to the ERA-Interim data used to nudge the model) qualitatively reproduces the drop in water entering the stratosphere as observed by HALOE.*

  Reply: It is a common misunderstanding, that a nudged model exactly reproduces the nudged data, which implies that it would act like a chemistry transport model (CTM). This is not the case, because the nudged model develops its own physical state on sub-synoptic scale. Nudging means adding a tendency to the model calculated tendency of selected prognostic variables in the simulation, which in our case is a Newtonian relaxation towards ERA-Interim reanalysis data. Nudging, however, still allows the model to develop its own subgrid-scale physics. Please note: water vapour is not nudged in our simulations and thus the hydrological cycle evolves freely, yet reacting to the additional weak forcing through nudging. This is different from ERA-Interim water vapour, which is assimilated by a 4D-Var data assimilation scheme. Therefore, we do not expect that nudged models provide the same results in detail as the data used for nudging. This also holds for the nudged QBO: all simulations are nudged with the same data (Singapore winds), in order to improve the timing of west and east phases. The resulting winds are, however, not the same in RC1SD and RC1, RC2, because the model still generates its own wind profiles!

- *I just can't see how you can reach the conclusion from your model calculations that ENSO via SST pattern is key to the problem, when the runs that are forced with observed SSTs (and even QBO!) completely fail to produce a drop around the year 2000. While suspecting ENSO/SSTs to be involved is completely reasonable, the challenge is to demonstrate that this is indeed the case, and your model results - along with other model results (e.g. SPARC CCMVal2) - fail to demonstrate this connection. The conclusion that "appropriate boundary conditions" are required is not helpful given that only your fully nudged run - where essentially every variable is set to prescribed values (P24913/L5ff) - gives the qualitatively correct result.*

  Reply: Thank you for this comment. We refer to your statement that the RC1 simulation (completely) fails to produce a drop in water vapour. We do not agree, that the drop is not simulated at all. The amplitude (of the decline around year 2000, what we call "drop") is too small and the period of low water vapour values is clearly too short. This is all the more interesting, because the temperature anomaly of the cold point (Fig. 5) is reproduced. This raised the question, if the characteristic of the drop is masked by a too low cold point temperature in RC1.

  In this context, it maybe have been overlooked that the simulation RC1SDNT also reproduces a smaller drop (and shorter recovery period). The only difference between the setups of RC1SDNT and RC1SD is that RC1SD includes the relaxation of the (vertically dependent) global mean temperature, whereas RC1SDNT does NOT include this. Note that this option is possible because the nudging is performed in spectral space, and "wave-0" (corresponding to the global mean) can be omitted easily. In other words, the additional nudging of wave-0 in RC1SD (compared to RC1SDNT) causes an additional temperature bias correction, whereas in RC1SDNT only the synoptic patterns are nudged. As we show by comparing the water vapour of RC1SD with that of RC1SDNT, the temperature bias is responsible for the reduced drop amplitude. Note further that the mean temperature bias (compared to ERA-Interim) of RC1 is even larger than that of RC1SDNT.

  Thus, as a further proof for the role of the temperature bias, we shifted (in our output data) the temperature anomaly at the cold point of RC1 to the mean cold point temperature of RC1SD (adding the difference in mean cold point temperature (RC1SD-RC1) to the RC1 cold point temperature). Then, we calculated the saturation water vapour values and found a similar drop (though not identical) as in the RC1SD simulation. We add a Figure with the respective saturation water vapour values to the revised manuscript (see also Figure in this reply). Apparently, the too low cold point temperatures contribute significantly to the reduced water vapour variability.

Although, this point was already discussed in the manuscript (Section 3), we accept that the explanation was not clear enough. We will improve it in the revised manuscript.

Another misunderstanding might be the definition of "drop". We do not explain the "period of low water vapour after 2000" with ENSO / SST and QBO coincidence, only the sudden drop (decline) of water vapour in 2000. As we conclude, such a sudden decline nearly always occurs after a strong ENSO event with concurrent phase change of QBO from west phase to east after La Niña. The recovery thereafter is not connected to ENSO.

- *Second, given that the fully nudged run (RC1SD) is presumably very similar to reanalysis data, it is not surprising that you get a drop - this has been known for a decade (see e.g. Figure 2 of Fueglistaler and Haynes (2005)); the challenge with the drop today is proper attribution to processes (see above), and accurate quantification and reproduction of the magnitude. As shown in Fueglistaler et al. (2013), all mode reconstructions using a wide range of available temperature data give what you also find: a drop, but the magnitude of pre/post 2000 is smaller than observed by HALOE; however, if one compares to SAGEII, the agreement is much better. The manuscript does not mention this conundrum, and is vague in terms of assessment of the success of the model result (first, it is noted that there is a "small" discrepancy, while later the discrepancy is quantified to be 50% - see comments below).*

  Reply: You refer to model simulations by Fueglistaler et al. (2013), already describing the water vapour drop. These are trajectory calculations based on ERA-Interim data. However, we use an Eulerian model with a hydrological cycle that develops freely under different boundary conditions (SSTs, GHGs, solar forcing).

  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.

According to the conundrum of discrepancies between different observation based datasets: an analysis of this is beyond the scope of the present study (although we are going to mention it in the revised manuscript). To confuse the situation even more, we add another (recent) dataset of Hegglin et al. (2014) to our Fig. 1. This shows that our simulated water vapour drop lies in fact within the range of at least two observational data sets.

Referring to the confusion with “small discrepancy” “and 50% difference”, this indeed needs clarification. In the original manuscript we refer to two observational data sets: HALOE/Aura-MLS and HALOE/MIPAS. The former is used to compare our water vapour anomaly as a near global mean. With the latter we performed an analysis with respect to the zonal mean drop characteristics dependent on latitude: drop date, length and amplitude. This seems to be inconsistent and contradictory, but it is not the case. We included a text in Sec. 2.2 of the revised manuscript to clarify this aspect and refer explicitly to the different data sets used.

Concerning your statement that the “challenge with the drop today is the proper attribution to processes” we clearly agree. Yet, we see a need to clarify the simulated difference in drop appearance of RC1SD and RC1. We explicitly state that our simulations are not the first choice to explain the effect of certain processes for the appearance of the water vapour drop. In that case detailed sensitivity studies would be more appropriate. But this is not the focus of the manuscript.

- *Abstract, line 11: You date the “start date” to the “early days of 2000”; Fueglistaler (2012) argue, based on their Figure 9, that the drop dates around October 2000; it would be helpful if you could comment. (See also my comment below for P24916/L28 that the text does not explain how you determine this date.)*

  Reply: We define as the “drop” the decline of water vapour, shown in the time evolution of the anomaly in our Figures 1 and 1A. This starts with the highest value (i.e., the “drop onset”) which defines our start date. From Figure 9 of Fueglistaler (2012), we assume that here, the minimum is used to define the “drop date”. In order to avoid confusion and since the date definition is irrelevant for the process, we will replace “early days of 2000” by “in 2000” in the revised manuscript.

- *Further: “We show that the driving forces ... are tropical sea surface temperatures ...” As stated above, I don’t think that your model results support this statement. Rather, your Figure 4 demonstrates the failure of the SST-based model runs. The question then is whether (i) the model fails to correctly reproduce the effect of SSTs on the TTL, or whether (ii) some other process is involved.*

Reply: We conclude from Figure 4 that the amplitude of the water vapour variability (including that of the drop) is underrepresented in the free running setup RC1. For RC2 we do not expect a correct timing at all, since the model is forced by simulated SSTs in that case.

A sudden decline in water water vapour anomaly is nevertheless visible also in RC1 at roughly the right time, however, with underrepresented magnitude, i.e. a too large minimum.

Figure 5 furthermore shows that the effect of SSTs on the TTL temperature anomaly is reproduced in RC1, i.e., the results are comparable between RC1 and RC1SD. This seems to be a contradiction: the TTL temperature anomaly is reproduced in RC1, but the water vapour anomaly is not. This apparent contradiction can be resolved by the non-linearity of the Clausius-Clapeyron relation: The amplitude of the water vapour anomaly does not only depend on the temperature anomaly, but also on its absolute value: RC1 simulates a significantly lower cold point temperature compared to RC1SD. As a consequence the water vapour variability is also lower, because of the temperature bias at the cold point – not because of its variability.

To demonstrate this, we added (from the output data) the difference of the mean cold point temperatures (RC1SD-RC1) to the RC1 cold point temperatures, keeping the variability of RC1 as it was simulated and calculate the corresponding saturation water vapour mixing ratio over ice. This result is shown in the attached Figure (which we will also include as new Figure in the revised manuscript), showing saturation water vapour mixing ratios of the temperature-shifted RC1 simulation closer to RC1SD.

Thus, in brief the answer to your question is: the model fails to correctly simulate the mean cold point temperature and this causes the absence of a "deep" drop. Nevertheless, we agree with the referee that other processes cannot be excluded.

We will describe this analysis in more detail in the revised manuscript.

- *P24911/L10: "This has become the big conundrum ..." suggest to reformulate.*

  Reply: We will reformulate it.

- *P24911/L13: "An increase in stratospheric water is expected..." This is a bold statement, not supported by any reference. I assume that your statement is not based on theoretical arguments, but on model results - in which case it would be fair to cite the papers (I assume that you think of CCMVal results, so please refer this work here).*

  Reply: We agree. Our statement indeed refers to model results. In the revised manuscript we will provide an appropriate reference.

- *P24912/L12: "Randel and Jensen (2013) state ..." I found this section unclear; are you saying here that your paper is to some extent a rebuttal* *to their statement concerning model results? Also, since you argue here that your model runs perform better than those referred to by Randel and Jensen (line 27: "... indicating that it is possible to ...") it would be useful if you could briefly list here what exactly is better in your model than in those that you compare to - "appropriate boundary conditions" (Line 28) is very vague. In any case, as already state above, I don't think that your results support your claim.*

Reply: With "appropriate boundary conditions" we refer to the observed (or reanalysed) SSTs **and** the nudging (of temperature, divergence, vorticity and the log of the surface pressure and the mean temperature (wave number zero in spectral space)). This is indeed misleading and will be clarified in the revised text.

As mentioned above, 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. Again: 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 as outlined above.

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

Last but not least, a free running simulation with simulated SSTs (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 augmented by the difference in the QBO signal between the different simulations (see Figure 14). 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.

Thus, to answer your question: Yes, we think our nudged (!) model simulations are performing better than our free running simulations. But this is obvious and to be expected, but not the point. The point is that we reveal some strong indications to why this is the case.

Similar analyses and evaluation of other model simulations are definitely required to corroborate our findings. This is, however, beyond scope for our present study.

- *P24914/L2: Please be more specific what "slightly nudged" means; reference to Jöckel et al. (2015) is not sufficient since the QBO is a key* *factor. Of particular importance here would be over which pressure levels you nudge the model (to the Singapore wind, I presume?).*

Reply: We will describe the QBO nudging procedure in more detail in the revised Section 2.1. We omit the word "slightly" as it is indeed vague and rather state explicitely the relaxation time of 58 days.

- *P24914/L25: It is not quite correct to state that previous studies focused mostly "on its absolute value"; see e.g. Fueglistaler and Haynes (2005; their Figure 2a); Fueglistaler (2012); Fueglistaler et al. (2013). Also, note that the focus on some period-average is not a deficit of the studies you quote here, but is due to the fact that the year 2000 drop is unusually long; and the long duration is - aside from the magnitude of the drop - the main reason we're interested in this event. None of the oscillations in the satellite record after the pre/post-2000 change comes even close.*

  Reply: Thank you for this hint. As it is written, it is indeed misleading. We only wanted to underline the differences in methodology of our study compared to previous studies. We will omit this text passage.

- *P24915/L14ff: "In Fig 1 we show that our RC1SD simulation is able to closely reproduce the water vapour fluctuations as observed ..." and "is in accordance with observed values." and later "... drop in 2000 is slightly underestimated (about -0.12ppmv),..". Later on you quantify the mismatch as 50. As pointed out above, your results are in line with previously reported results; the remaining problem is the exact magnitude.*

  Reply: This is an important point to clarify. In Figure 1 we compare our results with the HALOE/Aura-MLS data (and additionally also with the data set as published by Hegglin et al. 2014 in the revised manuscript). In contrast, the drop characteristics as described in Figure 3 are the combined HALOE/MIPAS water vapour series. Both observational based data sets have been derived with different methodologies. This will be clarified in the revised text in Section 2.2.

  Note that only our free running simulation (RC1) results are in line with previously reported results. With nudging (RC1SD and RC1SDNT) we are able to reproduce also the magnitude within the uncertainty of presently available observational data. All we do is to analyse the differences between nudged and free running simulations, yielding - in our opinion - strong indications to responsible processes.

- *P24915/L12ff: The temperature dependence of Clausius-Clapeyron indeed poses a challenge for water vapor amplitudes in the presence of a mean temperature bias; however one can address this problem by analysing the amplitudes in terms frost point temperature variations (see Fueglistaler et al. (2013) for a discussion of the impact of a mean temperature bias on H2O variations; their Figure 5b/para33 is for the annual cycle, but extension to inter-annual variability is simple.)*

Reply: As stated above, we will add a new Figure in which the saturation water vapour mixing ratios are shown, that correspond to the cold point temperatures of RC1SD, RC1 and RC1 with a shifted cold point temperature. With that we visualise how the cold point temperature bias effects the water vapour variability.

- *P24916/L28: Here you state that the observations have a larger drop by 50% in the tropics - whereas above (p24915/L23) you wrote "slightly underestimated". Also -please explain how exactly you determine the "drop date"; as noted above, we have argued that the drop occurs around October (Fueglistaler 2012; and follow-up papers)- please explain the difference.*

  Reply: This is redundant. See our replies above.

- *P24917/L19ff: It is rather confusing that your temperatures (Fig 5) seem to give a different picture than your water vapor (Fig 4); for example, in Figure 5 the black and red lines are reasonably similar, which cannot be said for Figure 4; please explain.*

  Reply: As outlined above, this is due to the non-linearity of the Clausius-Clapeyron relation. See our reply above.

- *P24918/L24: Statistics based on ad-hoc thresholds are generally not useful; and I am concerned that your analysis here (0.5ppmv for one model run, 0.2ppmv for another model run) falls into this category. Please show that this is not a concern here, or remove the analysis.*

  Reply: The thresholds have been only used to simplify the search of drop events with preceding ENSO events. Thus, the result of event identification counting is independent of the selected values. We could have also started with the ENSO index and search for drop events after La Niña events. The result is the same. Furthermore, we do not base any statistical analyses on this counting.

- *P24919/L2ff: "... eruption of Mt Pinatubo had a significant impact on temperature and water vapour ...". Please provide a reference for this statement; see also detailed discussion in Fueglistaler (2012), and Fueglistaler et al. (2014; ACP): Observations suggest that part of the aerosol warming tendency was offset by an increase in dynamical forcing of upwelling. Models generally have problems to reproduce this effect and therefore produce a massive moistening of the stratosphere - which is what you also find in your additional sensitivity run mentioned below on line 7.*

  Reply: The reference is Löffler et al., 2015, ACPD.[1] and will be added in the revised manuscript.

---

[1] Löffler, M., Brinkop, S., & Jöckel, P.: Impact of major volcanic eruptions on stratospheric water vapour, Atmospheric Chemistry and Physics Discussions, 15, 34 407–34 437, doi: 10.5194/acpd-15-34407-2015, URL http://www.atmos-chem-phys-discuss.net/15/34407/2015/ (2015)

Löffler et al. analysed two nearly identical nudged (without mean temperature!) simulations, where only one simulated the effect of volcanoes. The major volcanic eruption of El Chichon and Mount Pinatubo were represented in one simulation by prescribing zonally and monthly averaged values of the aerosol radiative properties. The main finding of the paper is, that stratospheric water vapour is increased after the eruptions, resulting from increased heating rates and the subsequent changes in stratospheric and tropopause temperatures in the tropics. Any effect of increased upwelling is already represented in the nudging tendencies in both simulations and should vanish in the differences of the two model simulations. Only the effect of the volcano on upwelling thus remains. We agree, that models generally produce a probable too moist stratosphere after the volcanic eruption. Therefore, to be on the safe side, we neglected this period in our study. Furthermore, we will shorten the respective text passage in the manuscript due to the suggestions of another referee (Mark Schoeberl) and omit the addional sensitivity run.

- *P24920/L26/Figure 9 Please be specific which equation and terms you use.*

  Reply: We will state the used formula of Pearson's correlation coefficient in a new Appendix B of the revised manuscript.

  The residual circulation has been calculated with the formula for the transformed Eulerian mean as by Holton (2004), their equation 10.16b, for the tropics (20N-20S). This information will be added to the revised manuscript.

- *P24921/L11f: Can you clarify - are you saying that nudging to ERA-Interim slows down the upwelling in the TTL? Or is this simply an artefact arising from a difference in the pressure level of the cold point tropopause in the free-running simulation relative to reality/nudged version?*

  Reply: True, upwelling is slower in the nudged simulation RC1SD (see also Jöckel et al., 2015, GMDD). The nudging basically affects the whole momentum budget (e.g. resolved wave amplitudes are nudged, that largely drive upwelling) so it is not surprising that upwelling is different in the free running versus the nudged simulation. The overall shift in the tropopause is also likely contributing to upwelling differences at a given height. We have not yet performed the detailed analysis to answer the question which factor contributes to the upwelling differences most. However, for the purpose of this study, the important point is that upwelling is different in the simulations, and thus water vapour is transported at a different speed.

  We will add a sentence to clarify this.

- *P24922/L9: You could test in your model calculations whether the subtropics are involved; if it's only speculation please omit.*

  Reply: We will omit the sentence in the revised manuscript.

- *P24923/L1ff: If I understood correctly, you said earlier that the QBO nudging is equal in all model runs - why then this difference here? At what level do you truncate the nudging?*

  Reply: Yes, the QBO nudging setup is equal in all runs presented. Nevertheless, the resulting winds are not the same in RC1SD, RC1, and RC2, because the model generates its own wind profiles depending on the model configuration and setup. Nudging is not prescribing! Only for RC1SD, where divergence and vorticity and the logarithm of the surface pressure are nudged, too, the wind profiles are close to those of ERA-interim. The lower edge of the nudging region is 90 hPa. The nudging setup will be described in more detail in the revised manuscript (see above).

- *P24923/L13ff: I could not quite follow your reasoning here. ENSO is related to surface temperature anomalies, so I don't understand what you mean by "under normal SST conditions the influence (of ENSO, I assume?) on upwelling is smaller." What do you mean by "normal"? Please explain.*

  Reply: What we want to say: The SST anomalies have a direct influence on the upwelling. Thus, the stronger the ENSO event the larger the impact on upwelling. This simple rule of thumb is, however, in some cases violated due to other processes. By "normal" we meant "undisturbed conditions", i.a. without ENSO. We will formulate better in the revised manuscript.

- *P24923/L13ff: Are you saying that you accept a time lag (between cause and effect) *varying* between 6 and 34 months? Please correct me if I misunderstood, but a scientific cause-effect relationship requires a well-defined time lag.*

  Reply: You are absolutely right with your comment! However, the El Niño is not the cause of the drop. What we wanted to explain is that a drop (or better a large amplitude in water vapour anomaly) occurs after a period beginning with an El Niño event, followed by one or two La Niñas. During ENSO we see a strong correlation between SST and upwelling, resulting in a large amplitude in water vapour anomaly after the La Niña decays. Because El Niños and La Niñas, respectively, develop all differently and also last differently long, we see the different time lags between the onset of the El Niño and the following drop. The period with anomalous SSTs starts with an El Niño, but important for a drop is La Niña.

  The text will be improved accordingly.

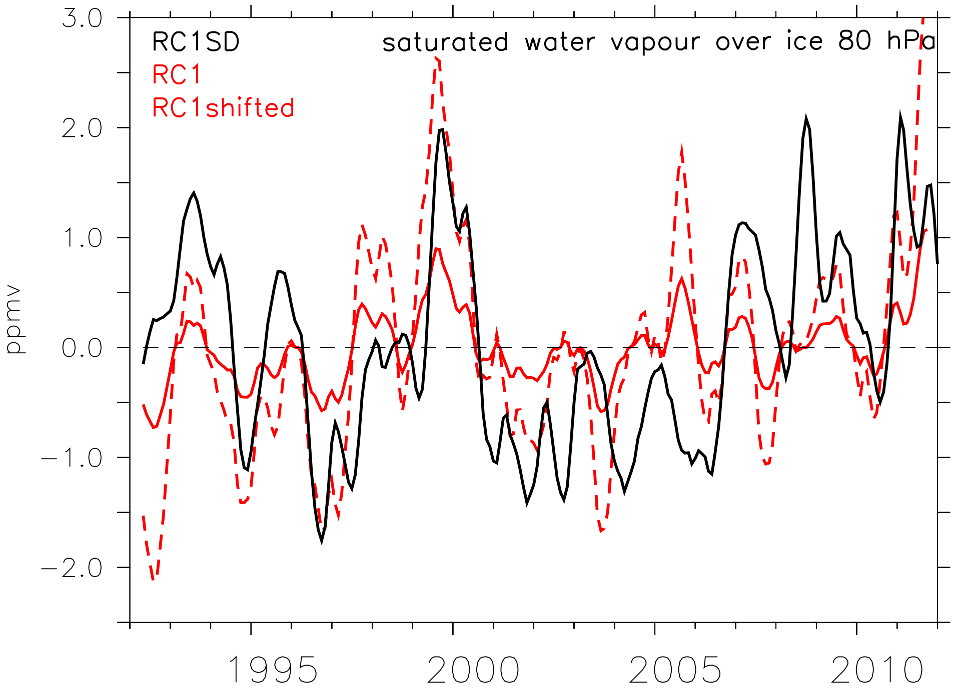

Figure 1: Saturation water vapour anomaly over ice (de-seasonalised, 6-month running mean) calculated from the respective cold point temperatures (10° S-10° N) of RC1SD and RC1 simulations. RC1shift: mean cold point temperature of RC1 is shifted to RC1SD mean cold point temperature. The mean cold point temperatures are: RC1SD: 192.1 K, and RC1: 186.0 K, RC1shift: 192.1 K)