Dear Editor

Thank you for editing our manuscript. In response to the reviewers' comments, we have re-written large segments of the manuscript. Specifically,

1. Much more attention is devoted to demonstrating the key role of Indian Ocean SSTs for the springtime tropical lower stratospheric response to El Nino.
2. Nonlinearity is now quantified.
3. Eight figures (out of the original sixteen) and their associated discussion have been removed in order to minimize distraction from the key points. Five of the eight now appear in supplemental material, and three have been eliminated entirely.
4. The discussion of the millennium drop has been shortened and focused.

The newly added text on the importance of the Indian Ocean is quite long, and for clarity it is copied below rather than within the detailed responses to the reviewers. The revised manuscript is ready for uploading, and we await your instructions.

Sincerely
Chaim Garfinkel (on behalf of the coauthors)

Why was the 97/98 El Niño tropospheric warming so distinct from other events? While this was the strongest El Niño over the period considered by this paper, the 1982/1983 El Niño was not much weaker than the 1997/1998 event as measured by the Nino3.4 index, yet the impact of the 1982/1983 on water vapor was qualitatively different. Furthermore, the upper tropospheric warming in the Central and East Pacific sectors for the 1982/1983 and 1997/1998 events (Figure 8) are similar. This suggests that the Central and East Pacific responses cannot explain the difference in stratospheric response. In contrast, these two events differed quite dramatically in the Indian Ocean (and more generally in zonally averaged tropical temperature). The 1997/1998 event led to remarkable impacts in the Indian Ocean: warm anomalies exceeded 2C locally over the West Indian Ocean and enhanced convection over Africa was anomalously strong even for EN (Webster et al., 1999; Su et al., 2001). Sea surface temperatures north of the equator were anomalously warm in the spring and summer of 1998 as well (Yu and Rienecker, 2000). As discussed in Garfinkel et al. (2013a), the cold point moves toward India in spring and thus warming in this area can impact water vapor. This difference in near surface conditions in the Indo-Pacific and Nino3.4 region is quantified in Figure 1. Conditions during the 1982/1983 event are shown with a red diamond, and during the 1997/1998 event with a large red x. Despite largely similar anomalies in the Nino3.4 region, the 1997/1998 event was characterized by remarkably warm anomalies in the Indo-Pacific that lie in the tail of the warming generated spontaneously in the coupled ocean-atmosphere model.

The importance of Indian Ocean SSTs for entry water vapor is quantified in Figure 9, which shows the regression coefficient between 85hPa water vapor and 2meter temperatures from 5S to 5N at each longitude grid point. We show both the regression coefficient in the annual average with no lag between water vapor and surface temperature and in springtime with 2meter temperatures leading water vapor by two months (Garfinkel et al., 2013a). The black curve shows the regression after linearly regressing out the BDC from the water vapor, and the blue curve regression after linearly regressing out the QBO from the water vapor.

In the annual average, warmer near-surface temperatures over the Central and Eastern Pacific lead to dehydration of the stratosphere in all three data sources, though during spring warming in the eastern Pacific leads to moistening of the stratosphere two months later. More importantly however, stratospheric water vapor is most sensitive to variability in the Indian Ocean basins and the Warm Pool region, with warmer temperatures in this region leading to enhanced water vapor in all three data sources in spring (and if the BDC influence on water vapor is regressed out, also in the annual average). Results are similar if correlations are examined (not shown).

Figure 10 demonstrates that the nonlinearity of the spring stratospheric response to EN is due to Indo-Pacific surface temperatures. It is constructed similarly to Figure 2, but years are stratified by 2meter temperatures from 50E to 150E instead of by the Nino3.4 index. Instead of the pronounced springtime nonlinearity evident in Figure 2, the lower stratospheric response to Indo-Pacific surface temperature is linear in all seasons. In winter, a warmer near surface leads to impacts similar to that of ENSO (compare top row of Figure 2 to 10). In March and April, on the other hand, a warmer Indo-Pacific leads to an accelerated BDC and a colder lower stratosphere, but to no robust changes in water vapor. In May and June, a warmer Indo-Pacific still leads to an accelerated BDC, but despite this accelerated BDC the lower stratosphere moistens. Results are similar for the AMIP integrations (not shown), with the 1997/1998 event leading to lower stratosphere moistening despite an accelerated BDC.

In summary, an ENSO event that more efficiently warms the mid-troposphere (such as the 1997/1998 event) by modifying SSTs in the Indian Ocean as well can more efficiently moisten the stratosphere. Strong EN events tend to have a stronger impact on the Indian Ocean than more moderate events (cf. Figure 1), and this tendency accounts for the nonlinearity in the impact of EN on the spring tropical lower stratosphere.

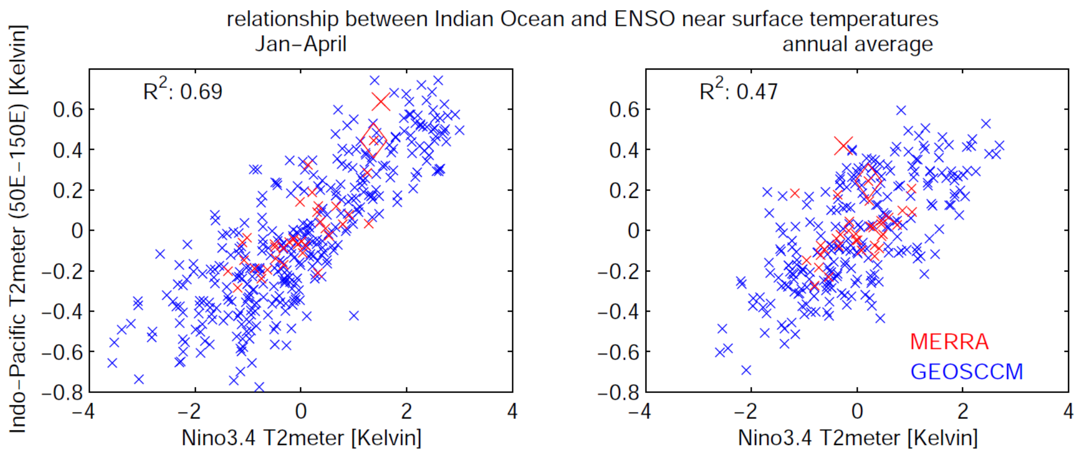


Figure 1. Relationship between near surface temperatures over the Nino3.4 region and over the Indian Ocean and Warm pool region from 5S to 5N in (blue) the GEOSCCM coupled ocean-atmosphere integration and in (red) MERRA reanalysis data. The EN event 1982/1983 is indicated with a large red diamond, and the EN event in 1997/1998 is indicated with a large x. (left) January through April; (right) annual average.

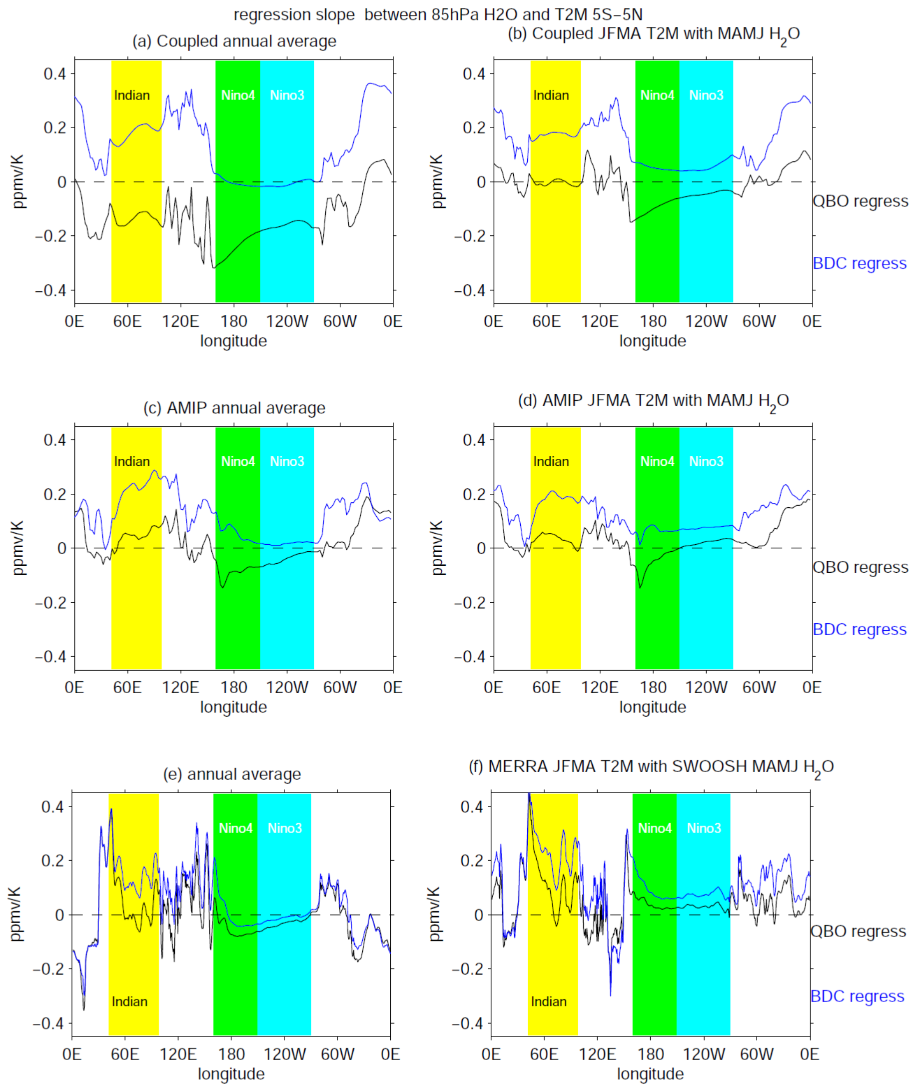


Figure 9. Regression coefficent between tropical (5S-5N) T2m and zonally averaged entry water vapor at 85hPa in (a-b) the last 240 years of a coupled ocean-atmosphere run; (c-d) the AMIP runs; (e-f) for SWOOSH water varpor and MERRA 2 meter temperatures. The longitude bands corresponding to the Indian Ocean, Nino3, and Nino4 regions are in color. The left column is for annual averaged quantities and the right column is for springtime (March through June) water vapor with T2m two months prior.

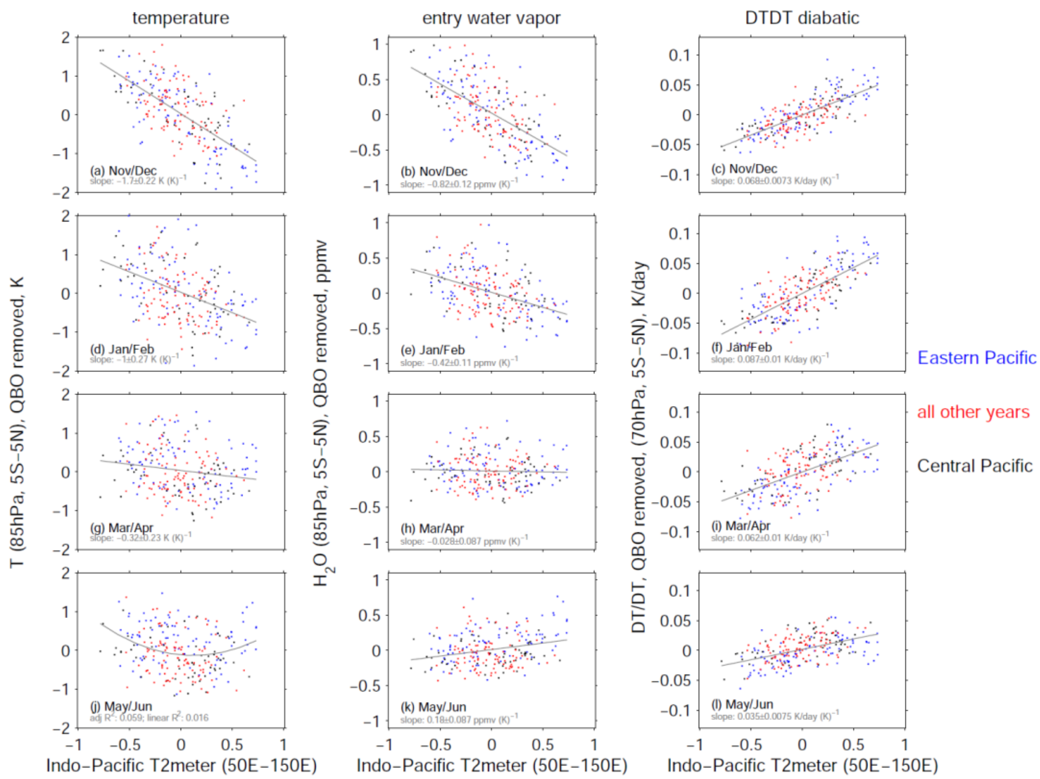


Figure 10. As in Figure 2 but stratifying years based on 2meter temperature from 50E to 150E, 5S-5N.

Reviewer #1

General comments: This paper discusses the impact of ENSO on the tropical lower-stratospheric (LS) temperature and water vapor by analyzing datasets composed of numerical simulations and reanalyses. The authors found that both La Nina (LN) and strong El Nino (EN) lead to wet stratosphere while moderate EN leads to dry strato-sphere even though the strength of stratospheric Brewer-Dobson circulation responds linearly to EN. The nonlinearity, i.e., the increase of ST water under strong EN condition is interpreted as the tropospheric warming extending up to the cold point that regulates the water entry to the stratosphere. The strong EN in 1997/98 and the following LN are attributed to the cause of the drop of ST water vapor in early 2001.

1. The analyses are limited to the temperature response and there found no argument on the modulation of pathways for the air entering the stratosphere. The coldness of the tropopause region does not necessarily result in the stratospheric dryness; as was pointed out by Bonazzola and Haynes (2004), “the sampling effect” as well as “the temperature effect” must be considered.

Thank you for pointing out this important effect. We agree that both sampling effects and temperature effects are important for TTL dehydration, and GEOSCCM of course includes both effects. However, we only have once-daily (daily averaged) output from the model on limited pressure levels and not on full model levels, and hence we are unable to quantify the sampling effect by running trajectories. We are therefore limited to analyzing temperature effects in our attempt to explain mechanistically why water vapor changes in the way it does in the GEOSCCM simulations.

Perhaps more importantly, there is quite a lot of scatter in our figures about the forced response, and this scatter represents (in part) the sampling effects to the final water vapor concentrations in the stratosphere. Stated another way, El Nino directly forces large scale changes in wind patterns and temperatures, and these effects are captured in the forced response as identified by the mean of the 42 ensemble members; the wind and temperature patterns in any specific integration among the 42 will differ from all of the others, and these deviations in the wind/temperature pattern are what we try to filter out by forming a large ensemble.

One could imagine that the sampling effects identified by Bonazzola and Haynes 2004 are due to such unforced variability that happened to be present in 1998 and 1999, and were not actually forced by the underlying ENSO event. However this hypothesis needs to be tested, and we don't mention this possibility in the text.

We have added the following to the data section: "The output necessary to run a trajectory model was not archived, and hence we cannot quantify the specific location of dehydration."

We have added the following to the conclusions in the list of unanswered questions: "Model output necessary to run a trajectory model was not archived, and hence we cannot directly address whether EN modifies the residence time in the coldest region of the tropical tropopause layer, an effect found to be important by Bonazzola and Haynes 2004. However, these sampling effects are included implicitly in GEOSCCM, and some of the diversity in response among the 42 ensemble members to an identical SST forcing is almost certaintly due to such sampling effects."

2. As for the millennial ST water drop, Fueglistaler (2012) and Hasebe and Noguchi (2016, ACP) identified its occurrence as October 2000 and September 2000, respectively. The current authors' mentioning of the year 2001 is different from these studies. Some arguments are required on the difference in the occurrence time and, most importantly, the driving mechanism.

Figures 14 and 15 in the initial submission were based on annual averaged values of water vapor. It is evident from figure 15 that individual integrations disagree about the specific year of the drop, and hence we don't expect GEOSCCM to capture the monthly variability in water vapor that was observed. Note that the QBO phase differs in GEOSCCM as compared to that observed, and the individual wave events driving the BDC differ as well.

We now clarify that these figures are based on annual timescales, and that the drop is fully consistent with the timing of the observed drop in these two publications. We also include a preamble that clarifies that the phase of ENSO from 1998 through 2004 was appropriate for a drop in late 2000.

3. It is not clear how “anomalies” and means are defined in many variables such as LS temperature, SST, heating rate etc. “Anomalous” labeled for vertical axes is not appropriate.

   We now define anomalies in the methods section as the deviation from the monthly climatology for each data set.

   We assume the reviewer is referring to “anomalous” on the vertical axis for figures 1,2, 3, 13-16, and this word has been removed.

4. The authors’ notion of nonlinearity is evident only in those shown in cyan with the Nino3.4 index greater than 2 (Figs. 1, 2, 9, and 10). Those points having the index > 2 will correspond to 1997/98 EP EN (p.6, l.30). In this context, it is important to study the features for this specific event in Section 5. The suggestion on the impact in the Indian Ocean is interesting, but there is no conclusive evidence having been shown.

Central Pacific events tend to be weaker (which the model captures) as the reviewer points out. However the regression coefficients (which use both types) are generally not sensitive to whether we exclude or include EP vs. CP events, though indeed there are a few cases where a linear regression would suffice for CP events but not for EP events. Hence we now note that the nonlinearity is less detectable for CP events both in the results section and in the conclusions.

More generally, we now quantify much more explicitly the importance of the Indian Ocean SSTs for the nonlinearity (see above). Specifically we have added three new figures and accompanying text, while removing some of the previous figures that are less important.

5. The argument in Section 6 is not convincing. The time series of H_2O and cold point temperature show large negative anomalies in 1997 followed by large positive anomalies in 1998. The H_2O drop is more pronounced (anomalies are larger in magnitude of negative values) in 1997 than in 2001, but there found no discussion on the cause of large drop in 1997. I don’t find any logical consequence in the statements given in page 9, lines 13 to 16.

In response to your comments and that of the second reviewer, we have rewritten the first half of section 6. At the beginning of section 6 we now clarify our expectations for the ability GEOSCCM to capture observed water vapor variability. Specifically, we note that the QBO and BDC in GEOSCCM does not match that in observations, and hence there should not be any specific expectation that the specific timing of drops should match that observed.

Rather, these experiments are useful for one purpose: quantifying the contribution of SSTs to the drop in 2000, and we limit our discussion to this point in the first half of section 6 (i.e. before we move onto 2015/2016). That the contribution is relatively minor (23%) is consistent with the previous work showing that other forcing factors (such as the BDC and QBO) are more important for entry water vapor variability. However other studies have suggested that SSTs play a role, but the literature lacks a clear quantification of their role.

In addition to these caveats that have now been added, we have rewritten and shortened the part of section 6 that relates to the millennium drop, in order to de-emphasize this relatively short analysis as compared to the rest of the paper.

6. Appendix: I don't understand why the mean age is discussed in the context of this paper. In addition, there found no explanation on how the mean age is estimated.

Mean age is calculated by a passive tracer whose concentration increases linearly with time. We now refer to Garfinkel et al 2017 where a full paragraph is devoted to details of this calculation.

We believe it is important to establish the linearity of the large scale BDC response, and hence we include this column. As we agree this is somewhat ancillary to the main point of this paper, it has been moved to supplemental material.

The bottom line will be that the LS water vapor tends to decrease in response to El Nino quantified by Nino 3.4 index but that the strong 1997/98 El Nino was exceptional in that it caused LS water increase. It is not clear if it is due to the warming in the TTL or it is related to the "flavor" (or type) of EP category it was classified. The mechanism has not been made clear by this analysis; it remains in the level of speculation. The terminology of "nonlinear response" may not be wrong, but it does not help understand the nature of LS water response to EN. The authors have interesting dataset obtained from ensemble runs, but it has not been analyzed satisfactorily. I recommend total rewrite of the manuscript after conducting analyses focusing on the specific features on 1997/98 EP EN. Considering the time necessary for the analysis, I suggest withdrawal of the present manuscript to consider re-submission.

We now better quantify the role of Indian Ocean SSTs for the water vapor response to ENSO, as we noted in the original submission that this feature was remarkable in the 97/98 event. Specifically three new figures have been added. This new analysis clarifies the source of the nonlinearity. This analysis is copied above.

Specific comments:

p.6, l.14-15: "EN leads to strong cooling" will be OK, but "LN leads to warming" is not obvious since the vertical axis is anomalies.

We now clarify "LN leads to warming relative to the climatology"

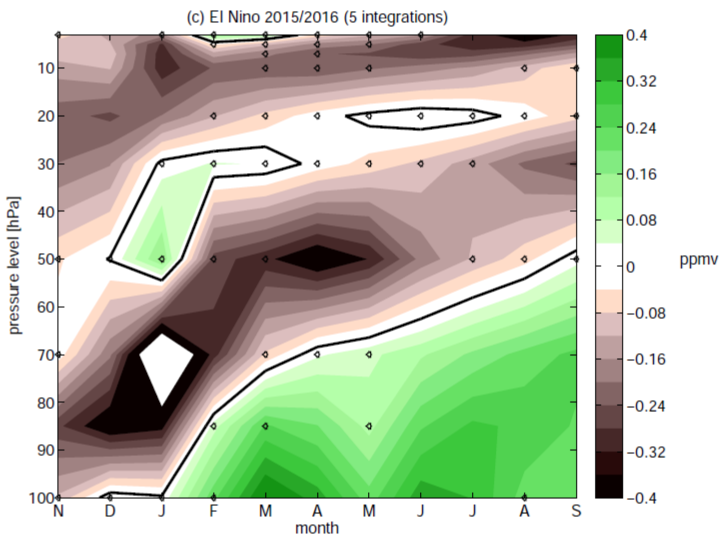
p.6, l.29: "This is especially evident in Figure 1il": In Figure 1i, the negative value of slope appears statistically significant, which may be interpreted as a linear response.

In response to reviewer 1, we have modified the way we display nonlinearity. As the slope is no longer displayed on this panel, this comment is not relevant in the revised manuscript.

To answer the reviewer more directly, the utility of a linear fit can be quantified using R^2 and adjusted R^2, and in this case the adjusted R^2 of a polynomial fit is substantially higher the R^2 of the linear fit.

Figure 4. The choice of green (+) and red (-) in color scale is confusing; the choice of the same color as in other figures (Figs. 5, 6 and so on) is recommended.

We agree that our original scale was confusing, and we now adapt a scale that we have seen used for precipitation in e.g. Kang and Polvani 2011. An example is below.



Figures 5 and 6: The distribution of cold region is only one aspect of TTL dehydration. There is no information on the location of the dehydration that is taking place for the air entering the stratosphere. The distribution of Lagrangian cold point was reported to have changed dramatically during 1997/98 El Nino (Fig. 8 of Hasebe and Noguchi, 2016).

As discussed above, we are unable to run a full trajectory model. However, the shift in the LCP in Hasebe and Noguchi appears consistent with the temperature changes shown in our figure 6. We now note this similarity and cite Hasebe and Noguchi in the text.

p.10, l.5-6: “regardless of their type”: It seems the nonlinearity appears only in EP type El Nino (2015/16 is also categorized as EP EN).

These words have been removed

Reviewer #2 (S. Fueglistaler)

Garfinkel et al. study the effect of ENSO on temperature around the tropical tropopause, and related to that on water entering the stratosphere. They study the problem using observations of stratospheric water vapor (the SWOOSH data set), tem-peratures from the MERRA reanalysis, and climate model simulations in a range of configurations. The paper's main point is that the response in tropopause temperature and water entering the stratosphere is non-linearly related to ENSO (as represented by some index) - such that both strong El Nino and La Nina lead to a temperature increase (and correspondingly moistening of the stratosphere). Using this result, they argue that the sequence of strong El Nino followed by strong La Ninas in the late 1990's led to elevated temperatures (and moister air) for a few years, which contributes to the 'drop' observed around September/October 2000. The hypothesis put forward is very interesting - but I have a number of concerns.

I could not help the impression that this paper was written up slightly careless. At times, the text reads more like a story than a scientific paper; similarly, the paper has problems finding the right tone. Consider the abstract. There, it is first written that: "The impact... is nonlinear." which leaves no room for doubt. However, the next sentence does not provide the hard evidence expected by the reader, but uses the rather weak word "appear" twice. Also, considering the seemingly straightforward hypothesis - the non-linearity of ENSO - I expected to be shown a plot that shows the non-linearity beyond doubt. Instead, the paper presents a full 16 figures that show a lot of information - most of it only qualitatively discussed. The paper would be much stronger if the authors were able to support the main point of their paper in one or two clearly drafted figures.

We agree that our original submission was a bit vague and qualitative. We have also removed three of the original figures as the discussion of them was (admittedly) overly hand-wavy, and they weren't truly necessary. Five other figures have been moved to supplemental material. The language in the abstract has been modified as well to be more confident. Finally, we have added several new figures that nail home the importance of Indian Ocean SSTs in a more quantitative way, as this is indeed the source of the nonlinearities. The revised version adopts a more scientific tone and hopefully is more convincing.

Figure 1 presumably presents the model data that best supports the argument for non-linearity-however the points are so small, and cyan has very little contrast, such that it is easy to overlook the datapoints supporting the hypothesis. Simply tweaking colors and symbol size would probably help.

Figure 1, coupled with the response to 97/98 shown in figure 2 (figure 2 and 3 in the revised manuscript), are indeed the figures that show the nonlinearity most clearly.

The marker sizes and colors have been changed in order to enhance readability. See figure 10 above (the third of the figures that relates to the importance of Indian Ocean SSTs) as an example.

Also, I'd like to see a more quantitative treatment of the non-linearity (i.e. it would be simply to compare the linear regression with a non-linear regression).

We now utilize the adjusted R-squared test, a standard technique in many disciplines to quantify the relative goodness of fits for a simple linear test and for a parabolic test.

Specifically, for all panels in figures 2,3, 4, 11, 13-16 we compute the adjusted R-squared using a polynomial fit and compare it to the R-squared for a simple linear fit. When the polynomial adjusted R-squared is more than 4/3 the linear R-squared, we adopt the polynomial fit. Note that in principle the 4/3 factor isn't necessary – if the adjusted R-squared is any higher than the linear R-squared then the polynomial fit is to be preferred. However we elect to be somewhat conservative and avoid over-fitting. For panels in which a polynomial fit is preferred, we show the adjusted R-squared for the polynomial fit and the linear R-squared. An example is in figure 10 above (the third of the figures that relates to the importance of Indian Ocean SSTs).

In all cases, the parabolic fit is better when we intuited it would be better in the original submission.

Also, it would be fair to show the statistical uncertainty in the "observational" data shown in Figure 3; we should be honest that the observational timeseries is really (too?) short to make statistically robust statements.

We aren't sure what exactly the reviewer intends as figure 3 (now figure 4) included a best fit line with 95% uncertainties. However we now note explicitly in the text that none of the regression lines for water vapor in SWOOSH are statistically significant.

The paper then applies the argument of non-linearity to explain the sudden and persistent drop of stratospheric water vapor around October 2000. However, there is a major conundrum pointed out in Fueglistaler et al. (J. Geophys Res., 2013) that needs to be addressed here: The arguably best representation of true temperatures in re-analysis data fails to properly produce a drop as observed in HALOE data. That is, the mechanism discussed in this paper applies to the large-scale effect of temperature and circulation, but the problem is that even if the free-running GCM would recover the reanalysis temperature perfectly, it would not be able to produce a drop as prominent as observed by HALOE. Correspondingly, it is not surprising that the drop diagnosed by the authors is only 23% of the HALOE drop. While there is plenty of good reason to have trust in HALOE data, it is crucial to note here that the stratospheric water vapor time series as observed by SAGEII agrees very well with the reanalysis-based model estimates (Fueglistaler et al., 2013). This needs to be discussed; and I would strongly encourage you to also consider quantifying the importance not against HALOE, but against the AMIP-mode model generated data (this helps your paper). However, the analysis of the drop as presented in Figure 11 is close to cherry-picking: anyone can see that what is labelled here as "decadal drop" is anything but a decadal drop. I also suspect that this time series does not compare favourably against SWOOSH at all - should this not be reason for serious concern given that this is an AMIP run?

We have significantly shortened the discussion of the water vapor millennium drop, as this is more of a secondary point of this paper than the main point. We also unfortunately set up unrealistic expectations as to what an AMIP model can actually accomplish as to capturing observed water vapor variability. We now clarify our expectations for the ability GEOSCCM to capture observed water vapor variability at the beginning of section 6. Specifically, we note that the QBO and BDC in GEOSCCM does not match that in observations, and hence there should not be any specific expectation that the specific timing and magnitude of the drop should match that observed.

Rather, these experiments are useful for one purpose: quantifying the contribution of SSTs to the drop in 2000, and we limit our discussion to this point in the first half of section 6 (i.e. before we move onto 2015/2016). That the contribution is relatively minor (approximately 23%) is consistent with previous work and the reviewer's intuition that other forcing factors (such as the BDC and QBO, or perhaps phenomena unrelated to cold point temperatures) are more important for entry water vapor variability near the end of 2000. However other studies have suggested that SSTs play a role, but the literature lacks a clear quantification of their role.

Quantifying the magnitude of the drop from a given satellite product is difficult due to missing data, and deciding among the many different ways to account for this missing data is beyond the scope of this paper. However, we agree that it important to note that there are discrepancies. Specifically, we have added that "the different satellite products underlying SWOOSH disagree as to the magnitude of the drop", and cite Fueglistaler et al 2013. We are now careful to write that **approximately** 23% of the observed drop is associated with SSTs, as the reviewer is correct that we do not know precisely how big the drop actually was.

To summarize, the paper by Garfinkel et al. touches many interesting points, and makes use of interesting numerical model runs. The paper needs, however, a major overhaul; the main points need to be worked out clearer in the data, and the discussion of the "drop" needs to be more careful. Given the recommendation for major revisions, I do not go further into the details of the current manuscript.