General comment:

The manuscript presents a new GOSAT retrieval for the estimation of the tropospheric water vapour isotopologue composition. The retrieval is very briefly presented and the data are briefly characterized. Furthermore, a brief empirical validation is performed by a comparison to water vapour isotopologue data produced from ground-based TCCON spectra.

The paper was indeed meant as an initial proof of concept, showing that water isotope retrievals from GOSAT are feasible and can complement other satellite retrievals which are more difficult to characterize.

Major comment:

Water isotopologue data obtained by space- or ground-based remote sensing techniques are potentially very useful for investigating the atmospheric water cycle, but respective measurements are very difficult. For this reason the remote sensing scientists' prime task should be to document the feasibility of the technique and the product quality in a detailed and extensive manner. In my opinion such documentation is a prerequisite for using the HDO/H2O data for scientific studies. In this context, I would like to encourage the authors to perform a more detailed documentation of the feasibility of their retrieval and of the quality of their GOSAT and TCCON HDO/H2O data products. I think that the near infrared HDO/H2O remote sensing experts should undertake similar efforts as their colleagues did for middle infrared water isotopologue retrievals: in the middle infrared HDO/H2O can be well detected (strong and well isolated H2O and HDO lines) and there has been a lot of efforts for documenting the quality of this product (e.g., Schneider et al., 2006; Worden et al., 2006; Schneider and Hase 2011; Worden et al., 2011; Lacour et al., 2012; Schneider et al., 2012).

We highly appreciate this comment and will go into more detail in the specific sections below. As the reviewer is aware of, another GOSAT HDO paper (by Hartmut Boesch et al) was submitted very shortly after our submission and is nicely complementary to our work as it outlines some of the more systematic retrieval errors in more detail.

Specific comments:

(1) Estimation of leading error sources:
I would very much appreciate if the authors provided information on the different error sources (measurement noise, modeling of scattering, spectroscopic issues, etc.):

- What are the leading error sources?

The primary error source is certainly measurement noise. Since deuterium depletion is a measure of relative changes, this error can be very large especially at low water vapor abundances. Other systematic error terms are now nicely explained in Boesch et al and will also hold for our paper. We will adequately refer to those in the revised version of the manuscript and extend the error description.

- Are there especially favorable measurement conditions?

Of course there are, i.e. clear sky bright scenes. In the error discussion as explained above, we will add a statement.

- Are there problematic measurement conditions?

High scattering layers (i.e. thin cirrus clouds) as explained in Boesch et al.

- How do atmospheric humidity levels affect the retrieved HDO/H2O amounts?

We will respond to this question in more detail later.

- Etc.
When presenting complex data like HDO/H2O it is very desirable to document that the sensitivity and the uncertainty of the product is theoretically well understood. This would give confidence in the dataset. Furthermore, a rough knowledge of the uncertainty levels can avoid mis-/over-interpretations and thus strongly increase the scientific value of the dataset.

We couldn't agree more with that statement.

(2) Constraint of HDO/H2O, page 6361, line 19:

In the current manuscript version you say that you perform the retrievals without any "H2O-HDO side constraints (besides the similarity in the a priori profiles)". What you describe here as a kind of secondary order constraint is implicit in your retrieval setup (hard constraint). I think it is a very strong constraint: (A) your kernels in Fig. 3 clearly show that the kernels of HDO and H2O are different, i.e. HDO-H2O constraint will definitively have an effect on your HDO/H2O result. (B) Near infrared spectra with a spectral resolution of 0.4cm-1 have a significant amount of information about the vertical distribution of water vapour (Schneider et al., 2011). Your constraint works against this information.

1) What we tried to convey is that both HDO and H2O columns are fitted entirely independently and we still believe that this is true. What you refer to is the fact that this can have implications on the retrieved HDO/H2O ratio, which can indeed be. However, if the end user is provided with the individual total column kernels of HDO and H2O independently, this will be taken into account. For the overall total column estimates, however, there is no cross-correlation between the retrievals which we

verified by looking at the full averaging kernel matrix. These matrices will also be provided to interested users, which should alleviate the concerns.
2) profile information: This is true for uplooking data with a well defined light-path. We have found, however, that near-infrared profile retrievals are susceptible to atmospheric scattering and that total column retrievals are much more robust to this error term.

(3) Sensitivity of your HDO/H2O retrieval, Fig. 3:
(A) Figure 3 shows that the sensitivities with respect to H2O and H DO are different. I think that this can cause significant artifacts in your retrieved HDO/H2O values and I would appreciate if you discussed this. For instance, imagine that there is a middle tropospheric increase of HDO and H2O, whereby HDO/H2O is not changed. However and according to Fig. 3, GOSAT will detect an increase in HDO/H2O (more sensitivity for HDO than for H2O).

Part of this was answered in the previous comment. The column kernels we display in the paper are not with respect to changes in VMR but to changes in (sub)column amounts. As the scale height of atmospheric water vapor is typically only 1-2km, changes from the prior in the free troposphere don't impact the final total column estimate strongly because these changes from the prior in terms of sub-column amounts are very small. For the layers contributing most to the total column, both kernels are very similar and close to unity. As we provide total column kernels to end users, this effect can easily be taken into account when doing model-measurement comparisons (as the scale height of water vapor also varies largely).

(B) Actually it is not sufficient to only look on the H2O and HDO kernels, if you want to document your sensitivity with respect to HDO/H2O. There are also crosscorrelations between the isotopologues.

Isotopically enabled atmospheric models typically calculate both HDO and H2O profiles and delta-D is a metric that is only derived from these model values. Our question is: Is a comparison against these models that apply the averaging kernels to both H2O and HDO column amounts first and then compute a AK-corrected delta-D comparison against GOSAT in any way worse than using the sophisticated methods used in Schneider et al. (2012)? We agree that this may be the case if there are severe cross-correlations in the retrieval itself but these cross-correlations in the column amount retrieval are not immediately apparent to us in near-infrared spectroscopy.

Furthermore, you might face the problem that the retrieved humidity and HDO/H2O states do not represent the same airmass, which makes an interpretation of the isotopologue data rather difficult.

As above, by providing sounding specific column kernels, the comparison against models should be unbiased.

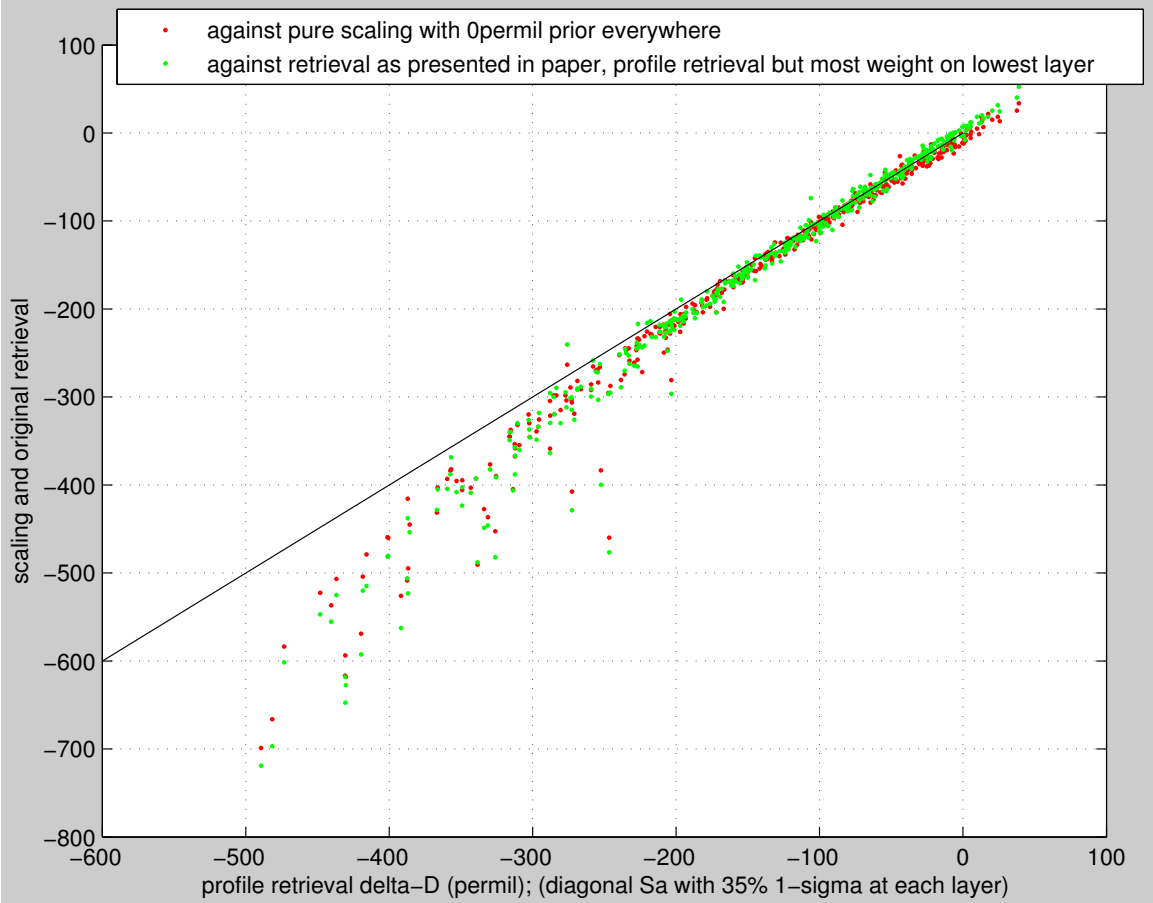
In Schneider et al. (2012) we show a procedure for documenting and correcting the crosscorrelations between humidity and HDO/H2O and for adjusting the different sensitivities of humidity and HDO/H2O. The attached Appendix shows an example of the humidity and HDO/H2O sensitivities and the cross-correlations between humidity and HDO/H2O. The documentation is made for two different water vapour isotopologue retrieval setups applying a typical IASI spectrum: first, for our Schneider and Hase (2011) IASI retrieval setup, and second for a setup that is very similar to your GOSAT setup (scaling of prescribed H2O and HDO profiles). I hope that this example can encourage you to analyze the HDO/H2O sensitivity of your retrieval in more detail.

We highly appreciate the reference to your work in AMTD and acknowledge the technical details of your study. However, we think that it is currently beyond the scope of this study.

In my opinion you should investigate
(1) to what extent your humidity and HDO/H2O product represent the same airmass (HDO/H2O is scientifically most useful if it is provided together with humidity), and
Strictly speaking, we don't really have a HDO/H2O product but a total column estimate of both HDO and H2O. In most cases (due to the scaling height of water vapor) both column estimates are sensitive to the same airmass (see Figure 3 between 1000 and 600hPa).

(2) to what extent your HDO/H2O product is affected by humidity interferences.
We are not sure we fully understand your point here but also conducted a small sensitivity study where we changed the retrieval setup for the Lamont GOSAT dataset. We compare three retrievals:

a) Our original retrieval (which was poorly documented as it is not a pure scaling retrieval but a profile retrieval with a purely diagonal Sa matrix (15% 1-sigma uncertainty for all layers but the lowest layer which comprised the bottom 10% of the total airmass and has a 1-sigma uncertainty of 1500%, we will document this more clearly in the revised version). In this case, the 10 retrieved HDO layers (equidistant in pressure) have the decreasing delta-D prior with height
b) Similar to a) but with 35% 1-sigma uncertainty ascribed to all layers, also the lowest one (i.e. overall more constrained but with less weighting towards the lowest layer). Prior profile as in a)
c) Pure total column scaling retrieval with a 0-permil a priori delta-D (strictly speaking the same HDO prior as H2O as delta-D is not part of our state vector).



The above figure illustrates the retrieval comparison with b) on the x-axis and a) and c) on the y-axis. The pure scaling retrieval as well as our original retrieval correspond very well with each other despite having a completely different HDO prior profile and retrieval technique. This speaks to the robustness of the retrieved column amounts. The main differences appear with respect to the more constrained profile retrieval at higher depletions (mostly because of the tougher constraint, not allowing too high depletions to be fitted). For the region between -200 and +20 (more like -275 and -70 in unbiased delta-D world) all retrieval agree to within the statistical noise.

(C) The TCCON retrieval is similar to your GOSAT retrieval and the TCCON water isotopologue product faces the same problems.

(4) Uncertainty in the ECMWF and NCEP profiles (used as a priori profiles), page 6366, line 5:

What about uncertainties in the ECMWF H2O data and uncertainties in the NCEP H2O data (used by the TCCON retrieval)? Can you assess the difference between the actual H2O and the H2O state described by the ECMWF/NCEP data state (e.g.,

by means of the numerous radiosondes launched at the Lamont site)? Please also consider that tropospheric water vapour varies on very small scales, i.e., the radiosonde and the remote sensing instrument very likely detect airmasses that have significantly different water vapour profiles. According to the sensitivity studies as shown in the Appendix an incorrect a priori profile assumption can significantly affect your retrieved HDO/H2O. Can you estimate the importance of this?

The radiosondes launched in Lamont are assimilated into the NCEP H2O product, so the a priori profile near local noon (there is an 11:30am local time sonde launch) is pretty close to the truth.

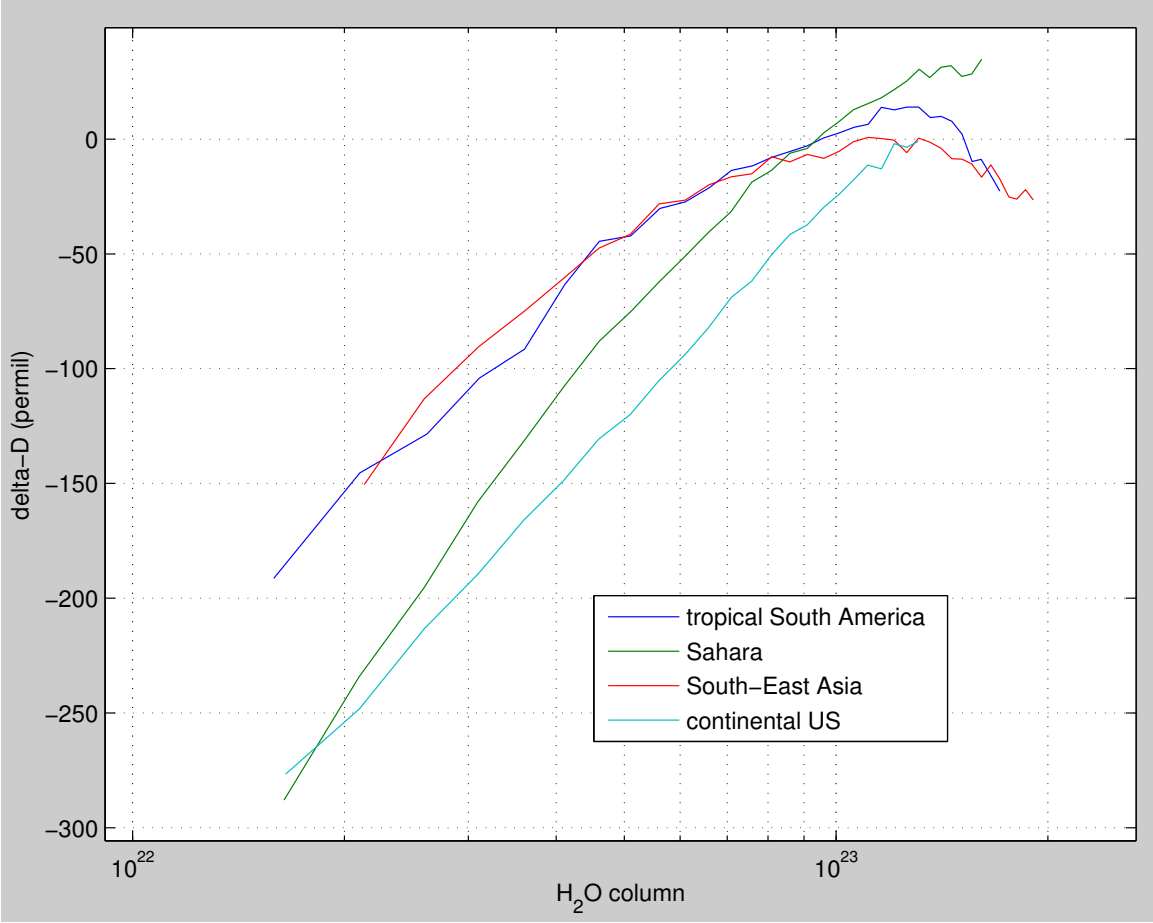
(5) Comparison between GOSAT and TCCON, page 6366, line 25:
I am not very happy that you compare two retrieval products, whose HDO/H2O sensitivities and uncertainties are not theoretically documented (see my comment (1) and (3)). Both the GOSAT and the TCCON retrieval setup apply similar a priori assumptions (e.g., a very similar HDO/H2O profile shape). At the same time it is not documented to what extent the retrieved HDO/H2O variability is, for instance, due to interferences from atmospheric humidity or uncertainties of the applied a priori profiles (ECMWF and NCEP for GOSAT and TCCON, respectively). I am a bit concerned that the GOSAT and the TCCON retrievals suffer from a common artifact. For instance, both the GOSAT and TCCON retrieval might suffer from similar humidity interferences on HDO/H2O. Then in both datasets the observed HDO/H2O variations would actually and to a large extent reflect real atmospheric humidity variations. I am not sure if your Figs. 4-7 really demonstrate that your HDO/H2O products provide additional information to H2O.

please see below

SUGGESTION: Please perform the following and straightforward test. Plot the TCCON and GOSAT data in a {H2O vs. HDO/H2O}-plot. Such plots can nicely document the added value of HDO/H2O measurements (e.g., Noone et al., 2011; Risi et al., 2012; Schneider et al., 2012). Then investigate similarities in HDO/H2O anomalies (deviations for the mean {H2O vs. HDO/H2O} correlation). If you can show that GOSAT and TCCON reveal the same anomalies your study would be much more convincing.

If we take the opening part of your comment 5 at face value, this wouldn't show us anything if both datasets have the same error sources?!
The plot you are interested is a little hard to be fully meaningful in our case (at least for the monthly means) because of a) the relatively high noise in the retrieval and b) the low deviations from the Rayleigh curve. Instead, we chose several different regions and computed the mean Rayleigh-type curve by binning values in H2O column amounts and then averaging these bins. Attached is a plot for the US, Sahara,

tropical South America and South East Asia. You clearly see the amount effect in tropical regions (which can certainly not be attributed to humidity interferences). You also see differences in the slope as well as the overall values at indentical H2O column amounts per region. This (to us) should resolve your worry that we don't really see "more" than just H2O variability.
We now added a section discussing this results and its comparison with the LMDZ model, showing that we actually observe some isotope physics.



(6) Comparison between GOSAT and SCIAMCHY, Figs. 8+9 and corresponding text: Here I have the same concern as for the GOSAT versus TCCON comparison. Can you assure that you really compare HDO/H2O and not just H2O (H2O variability might widely determine your commonly retrieved HDO/H2O signal).

I don't fully understand this point. True, the HDO/H2O signal depends on H2O through physical processes leading to the Rayleigh curve. However, the retrieval comparison should not determine whether information "on top of that" are discernible but (at least as a first step) whether the raw retrievals are comparable. Given that the time-periods of GOSAT and SCIAMACHY don't overlap (at least for HDO), your concern 6) is less of a worry than the potential inter-annual variability as mentioned by another reviewer. This first-off comparison was thus more meant

as a consistency (sanity?) check rather than a full-blown validation exercise for which the data is just not there yet.

(7) Why don't you compare to existing global datasets with assessed quality?
I wonder why you chose to validate your global GOSAT dataset at a single TCCON station instead of validating it with respect to the MUSICA dataset. Within the European project MUSICA water isotopologue data are available for ten globally distributed sites (Schneider et al., 2012). Of course these data are also not perfect, but at least their quality has already been extensively documented in several different theoretical and empirical assessment studies (a very complex work which is still ongoing). I think that a comparison to this "reference" would be more convincing than your comparison study.

This will certainly our next logical step. However, MUSICA measures in a completely different spectral region (1090-1330 cm-1 and 2650-3180 cm-1) as GOSAT. This is great for completely independent validation but since this is a first-pass comparison we think it's reasonable to look at ground measurements using a similar spectral region (to at least show that the impact of a widely different observing strategy yields similar results). As stated before, this paper was supposed to be a first step and proof of concept. A more extensive study was beyond the scope (and funding) of this manuscript. We did not in any way intend to disregard the great MUSICA dataset but will hopefully be able to continue this work and take all of these improvements into account.

Technical corrections:

Page 6361, line 17:
"chose" -> "chosen"

Done