REFEREE COMMENT:
This paper describes retrievals and associated error characterization of water vapor isotopologue measurements from a global ground-based FTIR network. The paper also describes some preliminary comparisons between these measurements and a global model. The paper is clearly written and the subject matter is highly appropriate for AMT. The authors have clearly stated the motivation for their work and the context under which it has been performed. They present a careful, considered and detailed error analysis, instilling confidence in a dataset that appears to offer promise for contributing to water cycle science.

REPLY:
Many thanks!

REFEREE COMMENT:
Nonetheless, I believe that there are areas where the manuscript could be improved. The term “interference from humidity” is used extensively throughout the manuscript. I think that this term is misleading. Interferences are generally considered to be things that one is not retrieving – here, the retrieval of humidity is an integral component of the approach. It is not really an interference. Perhaps the authors might consider the discussion in the context of “cross-state errors” instead?

REPLY:
Referee #1 had a very similar comment. We agree with you that the term “cross-state errors” would be more appropriate than the term “interference”.

REFEREE COMMENT:
Page 5370, lines 19-23: I think the authors ought to provide better justification of why their approach is superior to that proposed by Worden et al., 2006. The authors state that Worden et al.’s method uses “rather complex formulae” and is “not optimal”. Are the formulae really any more complicated than those presented here? In what sense is the Worden et al. approach not optimal? What, specifically, is better about the method here? Something about the way that the averaging kernels for the end product are supplied? Additional clarification of what the advantages of this method are over previous work would improve this paper.

REPLY:
Yes, we agree and we should better specify the advantages of our method.
We transform the {H2O, HDO}-state onto a {Humidity, δD}-proxy state (transformation matrix as given in Equation 6). In the transformed state we can treat the isotopologue products like any other trace gas products. In particular, we can perform analytic error estimations for isotopologue ratio remote sensing data in analogy to the method widely-used for trace gas remote sensing data (Rodgers, 2000). This is a large advantage and allows for characterizing the complex nature of the remote sensing isotopologue ratio data, using formalae well-known by the remote sensing community (for instance, compare our formulae (13) and (19) with Eq. (3.18) in Sect. 3.2.2 of Rodgers, 2000). Furthermore, we can provide averaging kernels. We think that these are advantage if compared to the method suggested by Worden et al. (2006). They calculate the errors by formulae as given in their Section 3.2. and provide H2O and HDO averaging kernels, but not δD kernels. Since the averaging kernels actually describe what is observed by the remote sensing system, we think that it is very important to provide δD kernels (or δD proxy kernels). Otherwise the δD remote sensing product might be easily misinterpreted.

REFEREE COMMENT:

There are a couple places where the citation of references could be more complete. Page 5363, line 10: "Schneider and Hase presented the first middle tropospheric optimal estimation retrieval. . ...using IASI. . ." What about Herbin et al.'s 2010 ACP paper? Page 5363, lines 4-5: Lossow et al. (the SMR reference) is cited, but the relevant MIPAS (e.g. Steinwagner et al., 2007; Payne et al., 2007; Nassar et al., 2007) are not. While this is technically covered by "references therein", it strikes me that it would be good practice to just cite references for retrievals from the other two instruments.

REPLY:

We will include the references you mentioned and also would like to explain why they are not in the current version of our manuscript:

The Lossow et al., (2011) paper "Comparison of HDO measurements from Envisat/MIPAS with observations by Odin/SMR and SCISAT/ACE-FTS" presents data from MIPAS, SMR, and ACE. It gives an overview of the different stratospheric and UT/LS water isotopologue satellite data. We thought that such overview for the stratospheric and UT/LS is enough, since the focus of our work is the troposphere (not the stratosphere and UT/LS).

Concerning the work of Herbin et al. (2010), please note that they estimate tropospheric H2O and HDO and not HDO/H2O. Such independent estimation of tropospheric H2O and HDO does generally not yield an appropriate tropospheric HDO/H2O estimate since the H2O and HDO kernels are rather different and tropospheric HDO and H2O distributions are very inhomogeneous. We can include the Herbin et al. (2010) reference explaining that it documents -- for the first time -- that IASI can detect different water vapour isotopologues.

REFEREE COMMENT:

Also, there are a couple of places where the language leans towards a proposal or sales pitch rather than a journal article. "PROFFIT introduces innovative retrieval options" (such as logarithmic retrievals). The capability to do logarithmic retrievals is certainly useful, but I am not sure I would call it innovative. Plenty other people have thought of doing that. "PROFFIT is currently the only retrieval code for ground-based remote sensing that supports an operational calculation of error Jacobian matrices." There are so many retrieval codes out there (presumably many that are applicable to ground-based remote sensing) that I find this a little hard to believe. Is it the operational part that is considered unique? What does it take to classify something as operational?

REPLY:

We will replace "PROFFIT introduces innovative retrieval options (such as logarithmic scale retrievals)" by "In addition PROFFIT allows for logarithmic scale retrievals and inter-species constraints".

Concerning the operational error Jacobian calculations our manuscript reads as follows: "PROFFIT is currently the only retrieval code for ground-based FTIR remote sensing that supports an operational calculation of error Jacobian matrices. This feature allows assignment of error bars to each individual measurement". Please note the specification on ground-based FTIR remote sensing. There are not so many ground-based FTIR remote sensing codes. We are aware of three different codes currently used by the ground-based FTIR community. One of them is PROFFIT and the other two do not operationally calculate error Jacobians. Due to the operational calculation of the error Jacobian matrices we can provide an uncertainty estimate for each individual data point. We think that providing consistent and comprehensive error bars for reference data is essential (we propose our FTIR data to be used as a reference for validating satellite sensor data and models). For this reason we think that it is rather important to mention this feature of our retrieval code.

REFEREE COMMENT:
I realize that the XCO2 retrievals are not the focus of this work, but I think the paper might benefit from some additional explanation in Section 5. For example, Page 5381, lines 22-25: "The de-seasonalised annual mean total CO2 column should be very similar at all the different sites around the globe." Can the authors quantify "very similar"? What about latitudinal gradients in CO2? Should these really be negligible in the de-seasonalised annual mean?

REPLY:
There is a latitudinal gradient in the de-seasonalised near-surface CO2 annual mean (anthropogenic sources of CO2 are mainly in the Northern Hemisphere). The difference is largest between the 30ºN-60ºN and the 60ºS-90ºS belt (5-6ppm, WDCGG-GAW 2012 summary, http://ds.data.jma.go.jp/gmd/wdcgg/products/summary/sum36/sum36.pdf, Figure 3.4a). Concerning XCO2, Reuter et al. (2012, http://www.atmos-meas-tech.net/5/1349/2012/, Figure 4) shows time series for different northern and southern hemispheric TCCON (Total Carbon Column Observing Network) stations. There are differences between northern and southern hemispheric de-seasonalised XCO2 annual means: for instance, in 2009 it is in average about 384 ppm at the southern hemispheric sites (Darwin, Wollongong and Lauder) and about 386-387 ppm at the northern hemispheric stations. For XCO2 the northern/southern hemispheric difference is about 2-3 ppm and apparently a bit smaller than for near-surface CO2. We think that the smaller latitudinal gradients in XCO2, if compared to near-surface CO2, can be explained by the following: first, the CO2 transport between the hemispheres occurs mainly in the free troposphere, i.e., inter-hemispheric transport affects XCO2 more directly than near-surface CO2 (CO2 emitted in the northern hemisphere arrives first to the southern hemispheric free troposphere and then it is transported downward through vertical mixing). Second, the northern hemispheric net emission (anthropogenic sources) as well as the southern hemispheric net absorption of CO2 (by the oceans) affects near-surface CO2 stronger than free tropospheric CO2 and XCO2.
Coming back to our Fig. 11, Wollongong, Lauder, and specially Arrival Heights (our southern hemispheric stations) should have by 2-3 ppm smaller XCO2 than the rest of stations. Indeed, we observe that Lauder and Arrival Heights have in average slightly smaller XCO2 than the other stations. We will briefly discuss this in the revised Section 5.

REFEREE COMMENT:
Figure 11: What exactly is meant by the 1 sigma scatter between stations?

REPLY:
"average 1σ scatter between different stations": we mean the standard deviation of de-seasonalised annual mean from different stations for a certain year. For instance, in 2000 the XCH4 values range from 378 to 381ppm (std of 1.9ppm, i.e., 5‰) and in 2004 from 387 to 388ppm (std of 0.4ppm, i.e., 1‰). We will explain this in the Figure caption.

REFEREE COMMENT:
In the conclusions, the authors state that the H2O profiles reflect real atmospheric variability between the lower and upper troposphere. What exactly is meant by this? Just that there are enough DOFS to distinguish between different altitude regions?

REPLY:
We mean that the ground-based FTIR systems are sensitive to water vapour variations from the lower up to the upper troposphere. However, we have to consider that the vertical resolution is rather modest (we achieve about 3 DOFS) and that we cannot resolve all the fine vertical structures of real tropospheric water vapour variability. This should be better

clarified in the conclusion section. Therefore, we will replace “they reflect real atmospheric variability between the lower and upper troposphere” by “the ground-based FTIR systems are sensitive to water vapour variations from the lower up to the upper troposphere”.