The paper presents mid-tropospheric δD observations obtained from thermal infrared nadir spectra measured by the IASI/METOP sensor. Since the IASI/METOP mission offers a unique combination of high spectra quality, long-term data availability, high spatial, and high temporal resolution, the sensor is very promising for providing long-term information about the tropospheric water vapour state. Such information is very useful for studying the atmospheric water cycle and for better understanding its future evolution. Since the water cycle is a key element of Earth's climate system such studies are important for improving the quality of climate projections. In this context, the paper is very suitable for publication in ACP.

However, I have two major comments and I would like to ask the authors to address them before publication:
- Since isotopologue ratio remote sensing is very demanding and in order to convince the reader of its feasibility, I think that the authors should pay more attention on a careful description and discussion of their retrieval methodology.
- I think it would be very helpful if the authors discussed their retrieval method in the context of the methods suggested by Worden et al. (2012) for TES and by Schneider and Hase (2011) for IASI/METOP.

**Specific comments:**

<u>Mention the difficulties of isotopologue ratio measurements already in the introduction section:</u> Isotopologue ratio remote sensing is very demanding. I think the additional difficulties of water vapour isotopologue ratio remote sensing – if compared to the remote sensing of atmospheric molecules – should be mentioned in the introduction section. I propose to add something like:
"… Tropospheric water vapour concentrations are very variable. Compared to this large variability the mid-tropospheric HDO/H2O ratio is rather stable. For measuring water vapour isotopologue ratios, we need a technique that is, firstly, sensitive over a large dynamic range, and secondly at the same time, very precise. Since it is very difficult for any measurement technique to optimally meet both requirements, tropospheric water vapour isotopologue ratio measurements are very difficult…"

<u>Page 13058, line 25ff:</u> "[…] it is anticipated that the retrieval will greatly depend on the choice of retrieval parameters, and in particular on the choice of the a priori information. For instance, Worden et al. (2012) show an improvement in the sensitivity of TES to δD after a change of the retrieval parameters".
Worden et al. (2012) improved the interpretation of the TES spectra by using a finer gridded model atmosphere and by fitting a broad spectral microwindow instead of small spectral microwindows. The same strategy has already been applied by Schneider and Hase (2011). However, the retrieval method of the authors does not apply these improvements: they use a sparse model atmosphere (limited to 10 levels below 10km) and rather small microwindows. I think this needs to be motivated. Why do the authors not build on the improvements introduced by Schneider and Hase (2011) and Worden et al. (2012)? I think they should briefly explain their motivation. For instance, if their main motivation is a very fast retrieval method they should mention it.

Section 3.2, model gridding: The authors limit the model atmosphere to the lowermost 10km of the atmosphere, use a gridding of 1km, and state that "variations of the state vector [at other altitudes] do not significantly affect the measurements".
As already mentioned in my previous comment, I think it needs to be explained why the authors do not build on the Schneider and Hase (2011) and Worden at al. (2012) experiences:
Schneider and Hase (2011) used a fine model grid and a model atmosphere up to the upper stratosphere for their IASI retrieval. Their $H_2O$ averaging kernels show still good sensitivity at 15km (sum along row of kernel matrix at 15km is still 0.5). Worden et al. (2012) found it important to increase the number of grid points for the new TES data version in order to increase the information content. Furthermore, in Fig. 3 the authors seem to indicate that there is still some information above 10km. I think that these discrepancies need to be discussed.

Section 3.2, applied spectral microwindow: The authors use very small spectral microwindows. This assures a high speed of the inversion algorithm. However, a small microwindows theoretically limits the sensitivity of the retrieval. Schneider and Hase (2011) use a broad microwindow. Worden et al. (2012) also changed to a broad microwindow for the new TES retrieval and argued that it is important, firstly to increase the sensitivity with respect to $H_2O$ and δD, and secondly to reduce interferences from $N_2O$ and $CH_4$. So the author's method of using a small microwindow might be very helpful to speed up the retrieval process but theoretically it will lead to reduced sensitivity and increased interference errors. In this context it would be very helpful if the authors discuss the information content that they obtain form their microwindows. I think it would be very important that they mention the degree of freedom of signal (DOF) they obtain for $H_2O$ and δD. They should discuss their values with the values obtained by Worden's TES retrieval and Schneider and Hase's IASI retrieval.

Section 3.2, cloud filtering: How was the cloud filtering performed? Please specify.
In case the authors use the EUMETSAT cloud flags, they should specify the flag used for excluding cloud contaminated observations.

Section 3.3, choice of a priori information, page 13061, line 24ff: "We found that the covariance matrix constructed in this way [applying LMDz model data] constrained the retrieval too much. This was evident by looking on the retrieval residuals y-F(x,b), which were found to be much larger than the instrumental noise. By overly constraining the retrieval, the retrieval remains too close to the a priori, and not all available information is extracted from the spectrum. […] $H_2O$ varies […] 260% at 7.5km.".
I find Fig. 2 not really informative. In my opinion here some details on the δD a priori data are needed. Instead of showing the covariance between HDO and $H_2O$ it would be much more interesting to know the covariance of δD (or ln[HDO]-ln[$H_2O$]). What is the covariance of ln[HDO]-ln[$H_2O$] you obtain from the model and what is the covariance finally used for the constraints, please give numbers! Maybe show a plot of the LMDz statistics: mean and standard deviation of ln[$H_2O$] and ln[HDO]-ln[HDO].
- Is the mean LMDz profile used as your a priori profile?
- Do you use a single/fixed a priori profile for all your IASI retrievals or is there some kind of latitudinal or seasonal dependence in your a priori? This should be clearly stated!

The approach of tuning the constraint in order to minimize the residual in the spectral fit can be dangerous. Part of the residuals above the noise level might be due to inconsistencies in the spectroscopic line parameters or due to an error in the applied temperature profiles. Addressing these residuals by reducing the δD constraint means that the retrieval misinterprets line parameter uncertainties or temperature errors as an atmospheric δD signal!
Furthermore, I am wondering if there might be a connection between the residuals you are talking about and your restricted model atmosphere. For instance, if you limit your model atmosphere to the first 10 km of the atmosphere, doesn't this mean that variability at higher altitudes is seen as residual in your simulated spectra?
Concerning H2O: To my knowledge a 260% variability for middle tropospheric H2O is rather large. At least I have never observed such high variabilities in radiosonde climatologies. I think the authors need to comment on that.

Section 3.4.3, error estimation, page 13065: The authors argue that Schneider and Hase (2011) found that the interference errors are not really contributing to the total error budget and that the same should apply for their retrieval. I think they should be a bit more careful: Schneider and Hase (2011) apply a broad spectral microwindow and retrieve the interference species simultaneously. The author's retrieval is different and the errors might also be different.

Validating measurements with a nudged model (LMDz): In principle a good idea, because there are few data available for validation. Applying the averaging kernels on the model data is important but at the same time dangerous, since it destroys the independency between the model and the measurement. For example: what if there is a large H2O interference on δD? It means that the retrieved δD is not independent from the retrieved H2O. Then both the retrieved δD and H2O will mainly reflect real atmospheric H2O variability (there will be no real information on δD). This interference will also be present in the averaging kernels. Consequently, the smoothed model data will well agree for δD, but what you actually compare is not δD, but H2O. I think the problems of comparing to model data, which are not independent to the observational data (because of the smoothing with the averaging kernels), should be briefly discussed.
Did you also compare to unsmoothed model data? This would not compare apples with apples but at least you would compare independent datasets. I think both comparisons are important: comparison to smoothed and unsmoothed model data.

Additional subsection in Section 3 for discussing the differences to other retrievals?:
Some of the aforementioned issues can be addressed within a subsection where the differences of the authors retrieval with the work of the Schneider and Hase (2011) are documented. Some examples of the differences (advantages/drawbacks) to Schneider and Hase that I quickly identified:
- The authors fit a rather small spectral microwindow. Advantage: short retrieval time; Drawback: less sensitivity?
- This constraint is different to Schneider and Hase. We assume a 100% variability for H2O and a 80‰ variability for δD for the whole troposphere (deduced from radiosonde climatology and Ehhalt et al., 1974, respectively).
- The Schneider and Hase retrieval provides data with much smaller errors, right?
- The authors present a retrieval that already works over land and ocean (Schneider and Hase so far only have presented retrievals for ocean scenes).

- The Schneider and Hase retrieval is empirically validated, while the authors compare their retrieval to model data.
- etc.

**Minor corrections:**

<u>Page 13069, line 19:</u> It is IsoGSM instead of IsoGCM.