This paper describes in detail an evaluation and error analysis of a water vapour profile retrieval algorithm (Multi-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water, MUSICA) that has been applied to the Infrared Atmospheric Soundings Interferometer (IASI) sensor flown onboard the EUMETSAT MetOp satellites. The algorithm is an optimal estimation (OE) algorithm after Rodgers (1990). The paper is well organized, thorough, precise and well written. As such, I recommend publication, but only after my comments/suggestions are addressed below:

We would like to thank the referee for their detailed review and valuable remarks and suggestions, which we address in the following.

General Comments

1. The Authors have neglected to reference any of the previous work regarding satellite sounder validation as detailed in numerous publications, especially those pertaining to the NASA Atmospheric Infrared Sounder (AIRS). In fact, I don't even recall that the AIRS sounder was even given a mention in this paper. As the Authors must know, acknowledgement of previous related work is a very important aspect of science publications. I can provide example publications, but I'm sure the Authors are already aware of them. It is important that this previous work be acknowledged in the revision.

In the introduction we give very few examples of satellite sensors that have been used for the remote sensing of atmospheric water vapour. Our idea was to give examples for sensors that use different wavelength regions (we give the examples of GOME, MODIS, TES, IASI, AMSU). We do not mention AIRS and would like to apologise for this. In the revised version we will expand the review and mention more sensors (discuss the importance of AIRS) and also discuss previous validation work.

2. On Page 7 the Authors indicate that they are using Level 2 IASI products as components in their a priori state. Unless the Authors are using independent channels from those used in the EUMETSAT L2 product, then this technically speaking would not be an a priori, which by definition is a "virtual measurement" not used prior to the retrieval. Also, this begs the question: Why not simply use the L2 H_2O product? The Authors should acknowledge and clarify these considerations, and it is particularly important that they explain the latter question — else the paper's novelty and/or utility comes into question.

We use the EUMETSAT L2 atmospheric temperature as the a priori temperature of our atmospheric temperature retrievals. We constrain our atmospheric temperature retrieval by a matrix that mimics the inverse of the EUMETSAT L2 atmospheric temperature uncertainty covariance matrix (see Sect. 2.2), i.e. for our atmospheric temperature solution state we allow for variations with respect to the EUMETSAT L2 atmospheric temperatures that are in accordance to the EUMETSAT L2 atmospheric temperature uncertainties. Our atmospheric temperature product is not independent from the EUMETSAT L2 atmospheric temperatures, instead it is an update of the EUMETSAT data obtained by fitting the spectral region used in our retrieval.
The skin temperature is freely fitted, i.e. there is no constraint and the EUMETSAT L2 skin temperature is used as the first guess for the iterative retrieval method.
Our H_2O (and HDO) a priori is a unique a priori for all the retrievals (no temporal and spatial variation, it is the same for all seasons and latitudes). This unique a priori is a mean profile from model calculations. In Sect 2.2 of the revised version we will give more details on the a priori information used by our retrieval.
The a priori data of the other species (N_2O, CH_4, HNO_3, CO_2) are also from model calculations and unique for all the retrievals (no temporal and spatial variation it is the same for all seasons and latitudes).

Specific Comments

• Page 2, Line 26: "MUSICA" — acronym should also be defined here.
Ok!

• Section 2.1: Nicely written introduction to the OE methodology — this is greatly appreciated.
Thanks!

• Page 3, Lines 3–4: This statement needs to be qualified with adequate references.
As the authors probably know, the NASA AIRS sounding algorithm (which is the pathfinder high spectral resolution infrared sounder) does not use the Rogers version of OE.
Ok, we will give references of satellite atmospheric remote sensing products generated by the optimal estimation method according to Rodgers (2000).

• Page 3, Equation (2): recommend that this be expressed as a formal equation — the “cost” term can be explicitly featured on the left side (a standard variable is J, but the authors are free to use whichever they choose).
Ok!

• Page 4, Lines 19–20: Acronym should be defined above at first occurrence.
Yes, right, we will define it in the introduction and not here.

• Page 4, Line 24: “HDO” — this chemical formula is not something I’ve encountered — although I (and astute readers) may deduce that it has something to do with the isotope formulas, the Authors should nevertheless define it and describe why it’s relevant and/or important to this paper.
In the revised manuscript we will better explain that the MUSICA IASI processor fits H2O and HDO (actually the HDO/H2O ratio). HDO and H2O vary not fully in parallel and when fitting high resolution spectra in the thermal infrared it is important to consider the different isotopologues. Furthermore, by performing the optimal estimation of H2O and HDO/H2O we can generate a {H2O,HDO/H2O} pair product that is very useful for investigating moisture transport processes (e.g. David Noone, Journal of Climate 2012, doi:10.1175/JCLI-D-11-00582.1).

• Page 5, Lines 2–3: By "single a priori" do you mean a single a priori profile globally? If so, where do you obtain this?
Yes we use a unique a priori for all the retrievals (no temporal and spatial variation, it is the same for all seasons and latitudes). This unique a priori is a mean profile from model calculations.

• Page 5, Line 5: “there is no constraint on the surface temperature” — What is meant by this? What is the a priori for surface temperature?
Surface skin temperature is freely fitted and we use the EUMETSAT L2 skin temperature as the first guess when starting the iterations.

• Page 5, Lines 14–15: Although not critical to the current paper, the Authors may wish to consider more recent models in future work, such as Watts et al. (1996), Masuda (2006), or the JCSDA CRTM model (Nalli et al., 2008), and perhaps acknowledge this in the text. The Masuda et al. (1988) model is known to have significant biases at larger scan angles. Also, what is being used as the surface wind speed?
We assume a wind speed of 5 m/s. A wind speed of 0 m/s or 15 m/s would change the emissivities in the fitted 1190-1400 cm-1 wavenumber region by up to 0.5% (the effect is most severe at the maximum satellite zenith angle, which is a bit smaller than 60° in the case of IASI). However, this 0.5% emissivity error would be almost uniform over the whole fitted wavenumber region and thus has an extremely weak effect on the retrieval products. We fit a 200cm-1 broad window and uniform emissivity errors will be largely corrected by the surface skin temperature fit. If the emissivity error was varying with frequency then it would have a significant effect on the retrieval products.
Many thanks for this comment. A comparison to other models can allow an estimation of the uncertainty in the Masuda (1988) data.

• Page 6, Lines 19–20: The meaning of this sentence is not clear — I’m not sure how different time periods at the different locations means that the the dataset is not uniform.
Sect. 3.2-3.5 and Table 1 inform about the differences for the different time periods. The retrievals at Manus Island and the Lindenberg retrievals for 2008 use IASI spectra measured after October 2007. Then IASI L2 products are available and we use the IASI L2 cloud product to identify clouds and the IASI L2 atmospheric temperature product as the a priori temperature in our atmospheric temperature retrieval. This is different for the Lindenberg and Sodankylä retrievals made with spectra measured in summer 2007. For this time period no IASI L2 products are available. We use the Vaisala radiosonde: (1) together with the cloud detection algorithm of Zhang et al. (2010) for identifying clouds and (2) for creating an a priori atmospheric temperature. Because a thermal nadir retrieval product depends on the assumed a priori

temperature and is also affected by clouds (see error estimations as summarized in Figs. 7 and 9, respectively), the usage of systematically different a priori temperatures and different cloud detection algorithms can cause a significant bias between the two time periods.

• Page 6, Section 3.1: What is the source for the Vaisala sensor information? This should also be included in the references.
We are not sure if we understand this comment, because we think that we provide the relevant information in the manuscript. At the end of Sect. 3.1 we write: "We work with Vaisala RS92 data that have been processed by the GRUAN lead centre (http://www.gruan.org). The GRUAN data processing assures that the obtained humidity, pressure and temperature profiles are well-calibrated and highly accurate (Dirksen et al., 2014; Sommer et al., 2016)".

• Page 7, Line 5: I was not aware that there was a GRUAN site at Manus Island. When did this come online and is it still in operation?
Manus Island has GRUAN processed data for 2012 and 2013, but is currently inactive (see https://www.gruan.org/network/sites/)

• Page 7, Lines 8–9: See General Comment #2 above.
Yes, we use the EUMETSAT L2 atmospheric temperature as the a priori for our atmospheric temperature retrieval. However, our H2O (and HDO) a priori is a mean profile from model calculations and it is unique for all the retrievals (no temporal and spatial variation it is the same for all seasons and latitudes).

• Page 7, Lines 13–15: Why is this detail not given for the Manus site?
Yes, we will also give respective information for the Manus collocation (there the collocation has been performed by EUMETSAT).

• Page 7, Lines 24–25: I'm not sure I fully understand — if the radiosondes are being used as truth, then how can they be used as the a priori?
For Lindenberg 2007 and Sodankylä 2007 we use the radiosonde atmospheric temperature as the a priori for our atmospheric temperature retrieval. We do not use a H2O (or HDO) a priori from a measurement. Instead the used H2O (and HDO) a priori is from model calculations and it is the same for all dates and locations (it is the yellow star in the correlation plot of Fig. 11).

• Page 8, Line 5: "this altitudes" should be "these altitudes"
Ok!

• Page 12, Line 25: "about every 10 m" — I'm not sure this is correct. My understanding is that the balloon ascent is about 5 m/s, and the Vaisala processed radiosonde reports every second.
Ok, we will double check.

• Page 12, Line 27: Insert "statistically" between "performed" and "in two steps", thus "is performed statistically in two steps"
We are not sure if "statistically" really captures the method. Isn't it more that we have to fulfill two conditions: (1) partial columns should keep the same and (2) the profile should not oscillate too much?

• Page 13, Lines 3–4: Replace "In order to get the in situ profile data that are comparable to the remote sensing data we have to smooth the" simply with "The" and insert "may be smoothed" after x_{GRUAN} and "according"
Ok!

• Page 14, Line 17: The logarithmic dependence of these formulations is not immediately apparent in the second lines of these equations, which simply denote relative values. I recommend inserting the intermediary mathematical step.
Ok. A similar remark has been made by referee #1. We will add the following explanation: "Because $\Delta \ln(x) = \frac{\Delta x}{x}$ we interpret the logarithmic scale difference between IASI and GRUAN as the relative difference (and use GRUAN data in the denominator)."

• Page 15, Line 4: Recommend using percents here so that it is clear that one is talking about relative values.
Ok, we will use “%” for all these values.

• Page 17, Line 7: “ca” should be “can”
Ok!

• Page 24, Figure 3 caption: More information is needed in the caption. The gray lines look like the actual averaging kernel matrix, so I’m assuming the colored lines are the row kernels? Also, what do the heights in the legend correspond to?
Yes we agree. All lines show row kernels. The coloured lines are the row kernels for the altitudes as given in the legend and the gray lines depict the rest of the row kernels. We think that highlighting selected altitudes is helpful for the readability of the Figure.

• Page 26, Figure 5 caption: Delete “Please”
Ok!