Review of ”Feasibility of a space-borne terahertz heterodyne spectrometer for atomic oxygen and temperature in the mesosphere and lower thermosphere” by Peder Bagge Hansen et al.

Overview

This paper is a feasibility study for retrieving atomic oxygen concentration and temperature from 2.1 THz and 4.7 THz atomic oxygen emission spectra measured by a limb-sounding heterodyne spectrometer from a satellite in polar orbit. The main idea of the study is generating simulated measurement spectra (based on temperature and atomic oxygen density from NRLMSIS 2.1 and winds from the HWM14 model) and retrieving temperature and atomic oxygen profiles from them. Atmosphere is not treated as spherically symmetric by the retrieval: three consecutive atmospheric scans are considered at a time, thus improving the retrieval of horizontal structures. Feasibility of retrieving line-of-sight winds from these spectra is also briefly discussed.

In general the paper is well-written, well-organised, methods and results are generally very clearly presented. The authors make a convincing case that the kind of retrieval they demonstrate here is feasible. My main points of criticism (see General/Major comments for more detail) are that the vertical resolution limits for this retrieval were not really explored (the retrieved data has much coarser vertical sampling than the simulated atmosphere scans), and that authors show very little data on retrieved winds. I would have also liked to see a few more comments on why some of the methods were chosen.

From the technical perspective, the quality of the manuscript is excellent. I have found very few typos, and made only a couple of small language and terminology suggestions.

General/major comments

1. **Vertical resolution.** The retrieval setup described in the manuscript seems to be aimed at delivering high vertical resolution: the vertical sampling of the measurements in a scan is 1 km of tangent altitude at the bottom of the altitude range, and the instrument has “1.0km field of view (FWHM) at the 100km tangential point for the 2.1THz receiver and 0.45km for the 4.7THz receiver”. However, the authors chose to represent their retrieved temperature and atomic oxygen concentration profiles with a set of B-splines, which have the knots spaced 5 km apart at the lower end of the altitude range, closely following the way atmospheric quantities are represented in the NRLMSIS model used as a source for the simulated data. I believe this representation (although seemingly perfectly valid mathematically) unnecessarily limits the performance of the retrieval. In particular, the vertical sampling of the retrieved quantities ends up being much coarser than the vertical sampling of the simulated measurements. Therefore, there is no way to tell if the proposed measurement scheme can deliver the high resolution at the lower altitudes it seems to be designed to achieve.

   I understand that the authors want their simulated measurements to be based on realistic data, and it can be difficult to obtain upper atmosphere data with highly realistic small vertical scale structures. However, one could simply resample NRLMSIS data on a denser vertical grid (or just use B-spline representation with more splines) and add some arbitrary small vertical scale structures to the reference temperature and atomic oxygen profiles just to see if they can be faithfully retrieved. I believe that a vertical resolution test with not-so-realistic data would be better than effectively not exploring the vertical resolution limits at all.

   This is particularly relevant in the light of some of the results presented in the manuscript. Figures 12 and 17 clearly show that 100 km to 125 km altitude region seems to be the most problematic, the retrieved profiles fluctuate around the "true" values much more than in any other altitude range. Furthermore, the authors say that "the majority of the retrieval uncertainties are concluded to stem from receiver noise and not parametrisation inefficiencies". If that is the case, and retrievals in this altitude range are inherently less stable even with 5 km vertical sampling (distance between B-spline knots), then is it really worthwhile to use 1 km vertical sampling for the measurements? Perhaps it would be better to reduce the scanning time and thus improve horizontal resolution (make ground track sampling denser) instead?

More generally, the authors seem to have taken great care to keep number of retrieved parameters to absolute minimum. While it is, generally speaking, a good practice, the ratio between the amount of retrieved parameters and the number of measurements is quite low in this study. This is not a problem in itself, but if the number of retrieved variables was kept so low because introducing any more lead to serious problems, maybe the dense altitude sampling used here is not bringing the expected benefits?

It would be good if the authors could at least add some comments explaining their choice of vertical sampling and the main limiting factors on the vertical resolution of the proposed instrument.

2. **Wind retrievals**

   (a) The authors state that retrieval of wind profiles is outside the scope of the study, and that one free parameter for Doppler shift was added to each simulated spectrum. I would argue, that since most of the radiance observed along each LOS comes from the vicinity of the tangent point, retrieving one LOS wind value per measurement is a rough approximation of a retrieval of a vertical wind profile (wind projection to LOS direction), just with a highly simplified wind treatment in radiative transfer calculations. It would therefore be good to show the retrieved wind values as a function of tangent point altitude. Even if the authors do not claim that they have retrieved the winds well, it is important to show the results in detail, since any problems with wind retrieval could have had a significant effect on the retrieval of other quantities. In other words, since winds were included into generation of simulated measurements, I think they are within the scope of the study, even if the authors only claim to have achieved good results for temperature and atomic oxygen retrievals.

   (b) As far as I understood from the general measurement setup, 2.1THz and 4.7THz measurements would be taken simultaneously for a given tangent point altitude within a given scan, and then modelled using the same LOS. However, the Doppler shifts for winds in these two measurements are retrieved as independent from one another. Is that correct? If yes, would it not be more logical to retrieve one wind value per LOS, from which the corresponding Doppler shifts could be derived? This would reflect the fact that both channels of the instrument should hopefully see the same wind?

3. **Non-linearity of inversion.** It is generally considered very desirable to keep inverse modelling problems as linear as possible. This both allows one to use many standard results of inverse modelling theory (Rodgers (2000) that the authors cite contains a lot of them), reduce computational costs with highly efficient linear algebra libraries and generally make retrievals more stable. Radiative transfer equations, sadly, are often non-linear, but most authors try not to introduce additional non-linearity beyond the forward model. Therefore, I find the equation (7) of this manuscript quite odd. It seems to me that one could have simply used additive correction terms instead of the multiplicative ones. I have briefly skimmed Livesey and Reed (2000), that the authors cite as an inspiration for their spherical asymmetry correction. That work does indeed introduce the idea of handling multiple adjacent vertical profiles together for improved retrieval performance, but standard linear algebra techniques seem to be used there. Could the authors

comment on why this approach was chosen?

Minor/specific comments

1. Lines 21-22: I am not quite sure what the authors mean by “[...] systematically analyses concentric atmospheric layers”. I would say that the main intrinsic advantage of limb observation geometry is high vertical resolution. Onion peeling, on the other hand, is just one of the methods (and perhaps the simplest one), to do a limb retrieval. In my opinion, the fact that onion peeling retrieves the parameters of the outer layers first, and uses them as the “truth” to retrieve the inner layers is more of a flaw than an advantage.

2. Line 43: I would replace “non-spherical symmetric atmosphere” with “non-spherically symmetric atmosphere”. More generally, the authors mention this lack of spherical symmetry quite a few times, but the first really clear explanation of what they mean is given in line 176: “ [. . .] variations in the vertical profiles across different longitudes and latitudes, i.e. the spherical asymmetry of the atmosphere [. . .]”. I think this could be made clearer by either giving this explanation right from the beginning (i.e. line 43), or using some other term, like “horizontal structure of the atmosphere” instead of “spherical asymmetry”. After line 43, I was left wondering whether the authors mean the horizontal structure, or do they mean that the oblateness of the Earth was accounted for in their calculations.

Minor typos and suggestions

Here is the list of minor suggestions and typos that I have noticed. I do not expect point-by-point answers for these.

1. Line 59: Replace “5.2MHz (FHWM)” with “5.2MHz (FWHM)”.

2. Line 184: “Without introduction of α, the retrieved profiles would correspond to some kind of averages over the regions traversed by the LOS.” It seems to me that the benefit of correcting for spherical asymmetry in general is described here, and not an advantage of the particular parametrisation of the horizontal structure (i.e. the α angle). It would be good to make this clear.

3. Line 238: “This retrieval resembles the example of a more accurate retrieval.” Do the authors simply mean that this retrieval *is* an example of a more accurate retrieval? If it only “resembles the example”, then where is that other example shown or described? The same applies for the similar statement in line 241.

4. Penultimate line of the Figure 13 caption: replace “corresponds” with “correspond”.