An et al., "Tomographic retrieval algorithm of OH concentration profiles using Double Spatial Heterodyne Spectrometers" presents a novel proposed approach to remotely sensed atmospheric OH concentrations using a pair of heterodyne spectrometers to record the necessary information to retrieve 3D OH fields in the stratosphere and mesosphere. Given the central role of OH as an oxidant in atmospheric chemistry, this method would certainly be a valuable addition to the fleet of atmospheric monitoring satellites in orbit.

However, I do not find that this paper communicates the proposed observing strategy well enough for a rigorous evaluation of its practicality. I therefore recommend significant revisions and a second round of review before the paper is accepted for final publication.

General concerns

- There are many points in this paper where the authors assert a value (e.g. for uncertainty) or conclusion (e.g. the tomographic retrieval is faster than the iterative retrievals) without providing a citation, clear chain of logic, or data to support those points. I list as many as I could find in the remainder of this document. These points need to be addressed so that readers can make their own assessment of these claims.
- What are the plans to test this algorithm in practice once the DSHS is operational. Balloon flights? Aircraft flights? Including this as a future direction would make clear what the next step is.

Specific comments

Sect. 2.2

Section 2.2 in this paper is intended to outline the observing strategy that this proposed instrument will follow in order to obtain spectroscopic information needed to infer the 3D OH fields. This is a key point that must be communicated clearly for the remote sensing community to evaluate whether this instrument is feasible. However I found this section extremely difficult to read.

The main problem is it seems the authors are exploring different observing strategies and rejecting ones that will not accomplish their goals. It is not clear that this is what they are doing, and it is not clear how they determined whether a particular strategy met their requirements, or even what those requirements are. I recommend this section be completely rewritten to improve its clarity, addressing the following points in order:

- First, clearly explain the criteria that a particular observing strategy must meet to be considered successful
- Second, explain how the strategies were evaluated to determine if they met these criteria
- Third, describe the *final* strategy that meets the necessary criteria

- Only then describe the alternate strategies considered and why they were rejected, making it absolutely clear that these are rejected strategies.

As an example of what makes this confusing, I point to lines 184–193:

> The satellite positions and observation geometries are given at the time of T1. **The satellite positions at the time of T2** need to calculate for realizing the three-dimensional segmentation of the target atmosphere. It indicates that there are two satellite positions where the scanning directions of T2-SHS2 and T1-SHS1 are orthogonal to each other in the intersection region at the time of T2 based on the theory mentioned above. Figure 4 shows that there is a tangent point which is defined as P1 when the DSHS detects the atmosphere by the three-dimensional limb mode. The intersection region along the lines of sight of SHS1 (the blue line in the Fig. 4) and SHS2 are the observed data at the time of T1. **There will be two lines of sight (the red lines in the Fig. 4)** where the scanning direction of SHS2 is orthogonal to the SHS1s' under the premise that the working altitude of the satellite is constant. The two lines of sight at the time of T2 (T2-1 and T2-2) and a single line of sight at the time of T1 form two intersection regions R1 and R2 which are symmetrical about P1 along the line of sight at the time of T1.

It is not clear from this section that T2-1 and T2-2 are hypothetical positions, and the bolded text makes it sound like these are actual times when the satellite will make measurements.

I understand that, to some extent, it is necessary to explore these hypothetical options for different measurement times to explain the criteria (e.g. line 194, "That means, the smaller distance between the P1 and R1 (or R2) is, the higher spatial resolution will be."). However, as written, it is unclear when hypothetical options are being considered and when the authors are describing the final option that is chosen.

Some more specific points of confusion:

- p. 5, l. 141: "The hierarchical detection of a series of observed radiances in the same target area comes true through the movement of the satellite platform." Does "movement" refer to the satellite flying along its orbital path, or actively rotating to point the spectrometers at the target area?
- p. 5, l. 139–140 vs. Fig. 3: The description of the two sensors as scanning along and across the satellite's orbit implies, to me at least, that the main axes of the telescopes are pointing parallel to the satellite's direction of flight, but Fig. 3 looks like the sensors are pointing to the right of the direction of flight.
  - This is exacerbated by Table 3, which seems to indicate that there are different T2 points for different altitudes in the profile. Addressing the first point about describing how the satellite moves will help with this.
- p. 6, l. 154–170: Since it's not clear how the satellite will move, it's very hard to understand what's going on here.

- How long will the spectrometers record at T1 and T2? Is it very briefly, or is the spectrometer recording as it sweeps across the target zone?
- Perhaps more generally, what is happening between T1 and T2? Is the satellite reorienting to get the appropriate pointing vector for the measurement at T2? Are the spectrometers on during this time?
- Do I understand right that the point of Fig. 3b is that one cannot use data collected at only one time to retrieve the target area? If so, why not? That is not explained.
- It's not obvious to me why the stratgy shown in Fig. 3c of using SHS1 at T1 and SHS2 at T2 is better than using both SHS1 and SHS2 at both T1 and T2. Surely having SHS1-T1 + SHS2-T1 + SHS1-T2 + SHS2-T2 would provide a better constraint on the OH fields?

- In Fig. 4, why are there red boxes (presumably representing the satellite) on both ends of the T2-x lines of sight but not the T1 line of sight?

Retrieval algorithm (sect. 4.2)

Given that the tomographic retrieval described in this section is the major advance in this paper, more detail on how the retrieval works is necessary to fully evaluate it. Specifically, in Fig. 16: what precisely is meant by "SHS1/SHS2 observed data at time of T1/T2"—radiance at a single wavelength, a few selected wavelengths, or all wavelengths measured by the spectrometer?

Presumably it is not a single wavelength because that would make it extremely difficult to disentangle the effects of the background radiance from that of OH fluorescence. But if it is at all measured wavelengths, that seems like it would result in an *extremely* high dimensionality to deal with in the interpolation.

Then, in the first paragraph of sect. 5.1, the authors state that there are four databases, one per season. How are transitions between seasons handled? If it's an abrupt switch (say May 31st uses the spring database and June 1st uses the summer database), won't that introduce discontinuities to the retrieved OH time series?

For Fig. 17, how does the inversion result compare with the true profile that was used to simulate the radiances? Similarly, on line 570, the sentence "The OH concentrations obtained by the LSUV retrieval algorithm are credible in the upper atmosphere," is extemely vague. What level of agreement do you consider "credible"? How large a sample size did you use to determine this?

Other major concerns

- Line 580: I don't quite understand "the undisturbed parts are taken as the actual parameters, and the results based on these parts are the true OH fluorescence emission radiance. Then, some certain amounts as the disturbances are applied to the three parts above. The OH fluorescence emission radiance based on these parts is the incorrect results." What are the "undisturbed parts"? Do you mean that you take the values of each of the atmospheric model, Doppler effect, and instrument calibration from the actual retrieval, and add some error to one at a time? That could be much clearer. How do you determine how much error to add? Do you do this for every OH concentration in the test set or a subset?

  - line 597: "...the amount of ozone profile for the disturbances is set as ±30%." Why 30%?

- Line 652: "We assumed the instrument calibration errors are ±5%..." Why 5%? Please provide either a citation or the line of reasoning used to arrive at this value.

- Line 657: "It is about 0.32% for the six-dimensional cubic spline interpolation of OH concentrations, date parameters, longitude, latitude, solar zenith angle and azimuth angle after research and related experiment." What experiement did you do to determine the 0.32% value?

- Line 741: "The speed of traditional iterative retrieval algorithms is slow and will cost a lot of time." Can you quantify this? How much faster is the tomographic algorithm for, say, a single target area or a day's worth of data?

Technical corrections

Points for clarification

- Lines 252–255: What is meant by "The SCIATRAN absorption spectrum database is completed based on the HITRAN 2014 database...." Does this mean that the forward model kept the HITRAN 2012 database, but added the UV absorptions from HITRAN 2014? Or does the forward model use HITRAN 2014 instead of HITRAN 2012? Or something else?

- In Eq. (10) where does $N(j)$ come from, if $N_g(j)$ is either from the OH database or the iterative update?

- Is Section 5.2.4 where we switch from discussion errors in the OH radiance to errors in the actual OH concentration? Please make it clearer that we are transitioning if so; the section title could be "Analysis of total errors in the OH concentrations" and add a first sentence like "We now move from considering errors in the OH radiances to errors in the retrieved concentrations."

- Line 704: "It is a great improvement to compare with the results of the LSUV retrieval algorithm which has a characteristic of iteration." What are the LSUV errors, or where in the paper are they listed? I found some of them later in the paragraph (line 715 and on), but why are they not in Table 7 for immediate comparison with the tomographic errors?

- Line 709: "The corresponding accuracy ranges of MLS OH concentrations products are also given, and its useful height ranges are from 23 to 81 km." Where are these values given?

Typographic comments

- line 142: "hierarchical detection," not sure what is meant by this.
- line 228: "The observed data received by the DSHS with an ultra-high spectrum resolution makes up by two parts in this research..." does this mean that the data recorded by the spectrometers will **be made up of** two parts?
- line 247: Citation for the Lifbase software needed
- line 248: Citation for the Bremen atmospheric model needed
- lines 236–275: there are several different ideas in one paragraph (the solar spectrum, the OH spectrum, the OH concentrations database). Please give each its own paragraph.
- Eq. (5): units?
- line 429: "However, it cannot be done to obtain OH concentrations in a converse way directly"—should "converse" be "inverse"?
- Around line 575: it seems like there is a shift from comparing the iteration and lookup table methods to the formal error analysis of the lookup table method. This should be more clearly separated.
- line 646: "The distributions of errors have the same tendency, but the relative errors of tomographic retrieval algorithm are smaller than LSUV retrieval algorithm." Does this mean that the distribution of errors due to the Doppler shift is similar between the LSUV and tomographic algorithm? As written, it sounds like the distribution of errors are similar to the relative errors.
- line 647: "These indicate that tomographic retrieval algorithm can obtain more accurate OH concentrations although the Doppler effect cannot avoid" The tomographic algorithm is more accurate than what? The LSUV algorithm? The Doppler effect cannot avoid what? Does this mean the Doppler effect cannot be avoided by either algorithm?
- line 701: "As the Table 7 indicates that the total errors of OH concentrations are increasing as the heights rise." Since the errors only increase with altitude at first, perhaps instead: "As Table 7 indicates, the total errors of OH concentrations **initially** increase with height..."