

# ***Interactive comment on “Intercomparison and Evaluation of Ground- and Satellite-Based Stratospheric Ozone and Temperature profiles above Observatoire Haute Provence during the Lidar Validation NDACC Experiment (LAVANDE)” by Robin Wing et al.***

**Anonymous Referee #1**

Received and published: 5 June 2020

This manuscript presents the results of a blind inter-comparison campaign that took place at the Observatoire de Haute Provence (OHP), France, a well-known long-term atmospheric composition monitoring station of NDACC. The results cover the ozone and temperature measurements of 3 lidars permanently deployed at OHP (LTA, LiO3S and LiO3T), the mobile lidar STROZ from NASA-GSFC deployed at OHP for the occasion, co-located ECC ozonesondes, nearby radiosondes, and coincident satellite measurements from Aura-MLS and SABER.

C1

These inter-comparison campaigns are essential to characterize the performance of the ground-based instruments, often considered “reference measurements” when validating satellite-borne instruments, and when long-term intercalibration between these satellite measurements is needed. This, together with the intrinsic value of the OHP timeseries themselves, makes the publication of these results in AMT highly relevant.

The take-home messages, as written in the abstract and conclusion, are clear and provide a good basis for reference in the future use of these datasets. Overall, the methods used are appropriate, but in several instances, a lack of clarity or rigor casts some doubts on the validity of some of the results, or more importantly, their interpretation. Two examples are 1) the loose/ambiguous reference to uncertainty and how it is used in the manuscript, and 2) the comparison of lidar and satellite uncertainty estimates and resulting conclusions. For this reason, major revisions are recommended before the manuscript can be considered for AMT publication. My comments and suggestions (major and minor points) are included below.

Major points:

1) Historically, the ozonesondes have typically been considered “independent” measurements. The ozone correction, as described here (i.e., using SAOZ), makes them dependent on the SAOZ measurement as well as the balloon blasting altitude. Recently, there has been a global effort for ozone sounding homogenization worldwide led by the SHADOZ community. Does this effort apply to the French ozonesonde program? Was the French ozonesonde team involved in this effort? Please clarify the role of the ozonesondes: Are they considered reference or just correlative measurements?

2) It is difficult to figure out what the authors refer to as “uncertainty”. For example, Page 17, line 9-10, “average measurement noise” is mentioned, and then in the same sentence “standard deviation of the ozone difference”. Are the authors referring to the combined uncertainty of the two measurements? If so, please use “combined uncertainty” instead of “standard deviation”.

C2

3) Please make the clear distinction between what is random, what is systematic, and how these two types of uncertainty components are treated in the various parts of the manuscript. For example, in Page 20, line 16, it is claimed that the “uncertainty estimates. . .are too optimistic”. Do these estimates account for systematic effects as well (total uncertainty?), or just the random component? If just random, it is not surprising that they do not match the r.m.s. differences, as r.m.s. will also reflect the presence of pseudo-systematic errors (e.g., alignment error for LiO3S, or aerosol interference for LTA).

4) It is not clear what MLS single profile uncertainty is. Please clarify. A single profile uncertainty is used. Aren't several MLS profiles used in the comparisons? Please clarify.

5) The authors' interpretation of Fig 21 is overstated and inconsistent with that of the previous figure. The two sentences starting with “So this disagreement. . .” on page 34, line 8-9 assume that because disagreement is found for all lidars, then all lidars are “wrong”. What if the source of the disagreement originates in SABER's underestimated uncertainty or a systematic error in the SABER profiles? If STROZ uncertainty is underestimated, why don't we see it in the comparison with MLS (Fig 20)?

6) Section 3.1.1: There is no evidence of vertical offset in figures 4 and 5. There is a difference in the shape of the peak between MLS and the others, but this does not seem to be the result of an altitude offset. For example, the MLS and GB ozone profiles are on top of each other at all altitudes above 24 km and all altitudes below 12 km. Also there is no evidence on these figures that SABER ozone/temp and MLS temp are shifted in altitude.

Minor points:

The title should read “Observatoire de Haute Provence Âž

Introduction: the authors should focus less on listing all the past campaigns and more

C3

on explaining the purpose of those campaigns and their outcome (include quantitative results as well as the main take-home messages from these campaigns.

Page 3, Line 3: Add “Aura”

Page 3, Line 29: Replace “off-line” by “non-absorbed”

Page 4, 2.0.2.: The few technical details in this instrument description section do not convey the right message. Please specify that tropospheric DIAL requires more absorbing wavelengths (stronger UV) to measure ozone at ppb levels rather than ppm levels, which is why the wavelengths are different from stratospheric DIAL. Also, specify that the initial 266 nm beam is spectrally shifted by the Raman cell to produce 289 and 316 nm.

Page 4, line 13: Remove “absolute”

Page 4, 2.0.3: As for paragraph 2.0.2., the choice of information included in this paragraph is somewhat arbitrary. There are other corrections applied to the signals to obtain the temperature profile (background noise, dead-time, molecular and particulate extinction). There is also a temperature initialization procedure at the top of the profile. I do not think the range-square correction should be mentioned without mentioning the other effects. I would recommend to add more details, or simply to mention that this is the backscatter temperature lidar technique, obtained by downward integration of atmospheric density (cf. Hauchecorne and Chanin).

Page 5, line 19: Is GPH converted to geometric altitude before it is used for comparison with lidar? Please specify.

Page 8, line 4: The impact of effective vertical resolution mentioned line 13 should be mentioned here

Page 8, sentence starting with “The increased spring time variance”: What is the purpose of this sentence? Is it supposed to introduce work further down in this manuscript? Please clarify.

C4

Page 9, line 22: The large percent differences between MLS and the other instruments is unlikely to be associated with MLS vertical resolution. It is mainly because the ozone peak and ozone minimum are registered at different altitudes. MLS is capable of identifying these sharp transitions. The main reason for the observed difference is most likely the spatiotemporal coincidence and atmospheric variability

Page 10, line 3-4: This is inconsistent with the figure. In fact, the best agreement is below 25 km

Page 10, line 8: I do not see any disagreement in the altitude. The peak is just smoother, and SABER actually reproduces well the ozone minimum right below the peak

Page 11, line 3: I do not think the addition of a new Raman channel will reduce the warm bias. Please rephrase

Page 13, line 5: Typo

Page 17, line 3: Please clarify. Does "uncertainty estimated by the retrievals" include only random components (photon noise), or is it the total uncertainty? How is this average computed?

Page 17, Equation 1: Define L and N

Page 20, line 8-12: Please define "MLS individual profile uncertainty"? Is that precision (random) or total uncertainty? Shouldn't a "campaign mean" of the individual uncertainties be considered instead of a single profile uncertainty? (just like it was done for GB instruments)

Correlation diagnostics (section 4, page 15 and section 5): What is the purpose of the correlation diagnostic? This diagnostic seems to introduce more confusion than clarification on the origins of the differences between the instruments. For example the authors state that the method is sensitive to the size of the averaging window, "drastically increasing or decreasing the amplitude of this peak". Please clarify or remove

C5

this part to keep the discussions of Figures 7 and 15.

Page 25, line 1: What is the a priori source? What is the altitude of initialization? Is it the same altitude for LTA and STROZ? LiO3S initialization is much lower. What a priori do they use? In order to investigate the STROZ warm bias at the top, could an alternate data processing be done using the same a priori at the same altitude? This would remove any bias associated with the tie-on procedure.

Page 27, line 8, "between 60 and 80 km": This sentence is misleading. If STROZ and LTA use the same a priori source (MSIS?), it is not surprising that the correlation increases as the profiles approach their tie-on altitude. This high correlation does not demonstrate instrument performance.

Page 30, line 7: "beyond statistical uncertainty": Systematic uncertainty components must be included in Figure 18, especially if they are not negligible, for example, uncertainty associated with temperature initialization and background noise correction in the mesosphere, and possibly dead-time correction uncertainty at the bottom of the profiles

Page 30, line 16, "...and also the temperature estimate of fig 10 of Leblanc et al.) are too optimistic..": I do not understand this sentence. Leblanc's figure 10 shows an example of uncertainty budget for a different lidar system (unrelated to LAVANDE), not including the impact of aerosols or misalignment. It is not surprising to find different results here, especially if the LTA profiles are impacted by aerosol and/or misalignment.

Section 5: There is no attempt to explain the NCEP differences (SSU?, AMSU?) Any published reference? Did the authors consider using MERRA-2?

Page 37, line 3, "Other sources of uncertainty": The authors are correct that other sources of uncertainty must probably be accounted for. But they should also discuss the possibility of optimizing the instrument set up (in this case alignment) so that errors are minimized and the introduction of additional uncertainty sources is less relevant.

C6

Figures:

Figures 2-3: are too small.

Figure 4: On the right panel, the differences between the instruments are not shown below 8-10 km. Please plot the differences. It would be good to uniformize the instrument short names throughout all figures and text. For example, sometimes, we see "OHP 532 nm", sometimes "LTA".

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-9, 2020.