Review of the manuscript number AMT-2018-240 submitted to Atmospheric Measurement Techniques and entitled "Validation of the TOLNet Lidars: The Southern California Ozone Observation Project (SCOOP)" by T. Leblanc et al.

General remarks.

The main purpose of this paper is a thorough review of the results obtained during a tropospheric ozone lidar intercomparison campaign (SCOOP) held in Southern California in summer 2016. The number of ozonesondes launched (18), the number of lidar involved (5) and the good weather conditions during SCOOP make this campaign very interesting to derive statistically significant results about QA/QC of tropospheric ozone lidar. The paper is well written (although a little bit lengthy) and the results are convincing. My recommendation is to publish the paper, but I suggest some minor improvements to make it more readable and avoid drowning out the main results with less important details. A discussion with the results of similar campaigns conducted in Huntsville (Kuang et al. 2011) or in Europe (Trickl 2015, Papayannis et al. 2015, ...) is also missing. One suggestion for improvement is to shorten the discussions on error modeling in section 4. For example, the paragraph on comparison with GLASS algorithm in section 4.2 could be reduced since it is not used for processing the SCOOP data. Information about the relative magnitude of errors related to ozone DIAL measurements already exists in the Leblanc et al. 2016 paper published in AMT.

Detailed remarks and questions:

p2. l.11 add the recent review paper on Tropospheric Ozone in the TOAR Elementa special issue e.g. Gaudel et al. Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation. Elem Sci Anth. 2018;6(1):39. DOI: http://doi.org/10.1525/elementa.291

p.3 l.28 Merge Table 1 and 2 and put the telescope size below the laser characteristic to check if a large telescope balances (or not) a low laser energy. What is Power ON/OFF ? Explain the meaning of this parameter in caption.

p.5. l.17 The following sentence is not clear "*The background subtraction value is determined from the last ~2 km in range*" Do you mean 13-15 km ? Is it high enough ?

p.10 l.20 Fig.2 is quite nice. Some comments might be already inserted in the text to discuss the negative bias in the lower troposphere below 3.5 km. Why not using the AMOLITE record which covers the full period ?

p.10 l.22 Daytime upper range is below 8 km according to Fig.2. Change 8-9 km by 8 km.

p.15 l.15 It is a very good idea to apply a vertical smoothing on the ozonesonde. Table 6 shows different vertical resolutions from the 5 lidars, did you consider this variability as well ?

p.12 l.3 Statistical analysis of one-to-one comparison is obviously better than the comparison of the means of each instrument. Is it is really necessary to include Fig.5 ? Fig. 4 is already a good illustration of it.

p.13 l.13 Give a reference where the GLASS software is described. After reading section 4, I wonder why comparisons with GLASS are so much emphasized while at the end this software is not used for the processing of each lidar. Comparisons of the output of in-house lidar data processing using a simulated signal are probably good enough to make the point that errors related to the processing software is not the reason for discrepancies between the 5 lidars.

p.14 l.22 “with little impact from aerosol” Is it really true ? If yes provide a reference. Do you expect the same sensitivity to aerosol interference for the 5 lidars ?

p.15 l.4 Fig. 8 is very busy and errors related to molecular extinction correction or air density are known to be always smaller than the other terms (see Leblanc et al. 2016). Discussion of errors related to zero correction, random noise and counter saturation with ozone absoption cross-section error is good enough. Indeed the latter is a kind of reference to the intrinsic limitation of the DIAL technique.

p.16 l.24-25 There is also a significant bias at 10 km for TMTOL.

p.16 l.27 “a +10% bias at 3 km for TMTOL”. Better to say below 4 km according to panel (e) of Fig. 10. Bias below 4 km are different for the two systems with a tunable laser (LMOL and TOPAZ) and the other 3 using the Nd-Yag 4th harmonic and Raman cell. Negative bias for the systems without a Raman cell and positive bias for the others. It would be interesting to discuss this point. Any explanation ?

p.17 l.8 “a poor sampling statistic”. What do you mean ? Number of comparisons ?

p.16-17 section 5. This is a key section of the paper. So I think some features may be discussed a bit more: (1) the 5 lidar have the same positive bias of 8% at 7 km. Why ? Aerosol layers ? (2) TMTOL looks very noisy during the day while it is the more powerful lidar see table 1 and 2 (If it is related to a gain amplifier problem as discussed in section 5.2 may be good to mention is here as well) (3) the choice of the shorter off-wavelength for LMOL and TOPAZ probably balance the lower laser average power during the day (might be useful to mention this).

p.17 l.16 Fig.12 is quite difficult to read especially the respective contribution of random and systematic error. Instead of using the ozone profiles in the two left panel of Fig.12, it is more useful to show the measured differences shown in Fig. 11 in order to see if large differences match high error estimate. It also means that Fig. 11 and 12 could be merged.

p.17 l.31 “*The peak at 8 km comes from using a pair of very high-intensity channels, therefore leading to a higher estimate of saturation correction uncertainty (blue dash curve).*” There is no corresponding ozone difference in Fig. 11 for TMTOL during the night. Do it mean the uncertainty is overestimated ? In fact it is a general question for all lidars the large uncertainties are not always where the ozone difference is large. Why ? On the contrary, a large ozone difference in Fig. 11 does not always correspond to a large lidar uncertainty, e.g. for AMOLITE at 6.5 km during the day .

p.18 l.20 What is the reason for the optimization of AMOLITE, LMOL, TMTOL lidar data processing afterwards ? It would be useful to relate the improvements made to the need of corrections identified in the previous section. Effective vertical resolution was not mention as a problem in the previous section. Right it is hard to see the added value of this final improvement in the SCOOP final results.

p.19 l.1 the word “outstanding” is probably too strong in line 1 although Table 6 is certainly a very nice contribution to lidar QA/QC. Indeed the overall mean bias and RMS must used with care as it implies that several lidars must run together to reach this level of uncertainty. May be better to say that it show that such an overall mean bias of 0.7 ppbv and RMS of 1.6 ppbv are achievable by a tropospheric ozone lidar even if Table 6 shows sometimes larger bias or RMS for a single system.

p.19 Summary. Add one sentence about the TOLNET lidar QA/QC assessment with results from previous intercomparison campaigns.

p. 20 Conclusions. Discussion on future improvement mainly focuses on the unattended operation. It is surprising considering the results of SCOOP below 3 km that nothing is said to the need for more accurate measurements in the lowermost troposphere. Lidar contribution to air quality assessment projects or even satellite validation will be improved with better ozone vertical profiles below 3 km.