Review of" Inter-Comparison of Thermal-Optical Carbon Measurements by Sunset and DRI Analyzers Using the IMPROVE_A Protocol" by Zhang et al.

The manuscript (amt-2020-436) reported an inter-instrument comparison between two pairs of carbon analyzers using two large datasets from CSN. The reviewer fully agreed that these comparisons are valuable since inter-instrument comparisons remain limited in the literature. The authors present substantial analyses, but some analyses seem incomplete. The reviewer feels that the data analysis and interpretation can be improved, otherwise, the value of this work could be weakened. The following issues should be addressed before consideration of publication.

1) Line 65-69. Please elaborate on why the carbon analyzer was changed from DRI2015 to Sunset 5L in 2018? The transition from DRI2001 to DRI2015 in 2016 is easy to understand, but the 2018 instrument change was too close to the 2016 instrument change. The readers would wonder about the motivation behind the 2018 instrument transition.
2) Section 2.1 was well written. This section provides a useful description of the three instruments, which covered all key features of instruments from the engineering perspective. That is very helpful and vital for the readers to understand the differences in instrument design.
3) Line 180. The author stated that the mass loading of DIR2001 data was back-calculated from the air concentration data using a constant sample volume (33 m^3). That would introduce unnecessary uncertainties for the instrument comparison since the actual sample volume varied by samples. The deviation of sample volume could be small, but using a constant sample volume for back-calculation is not scientifically sound. The author should use the actual sample volume for the back-calculation of mass loading of DIR2001 data.
4) The reviewer suggests comparing the average temperature profile between DRI and Sunset instruments for the two sets of data. First, there could be differences in the actual temperature for a specific temperature plateau when implementing the same TOA protocol. In addition, since temperature ramping rates were not defined in most TOA protocols, the differences in temperature ramping rates between the three carbon analyzers could be one source of carbon fraction discrepancy and are worth further investigation.
5) Line 380-381. The low laser transmittance (T) signal after analysis indicates the remaining strong light-absorbing materials on the filter. These materials are refractory, but not necessarily to be EC. One possibility is metal oxides as shown in the figure below. The authors are encouraged to check the elements data to see if there are any correlations between crustal elements abundance (e.g. Fe) and the abnormal low laser T signal after analysis. If yes, that would be an indication of metal oxide's influence.

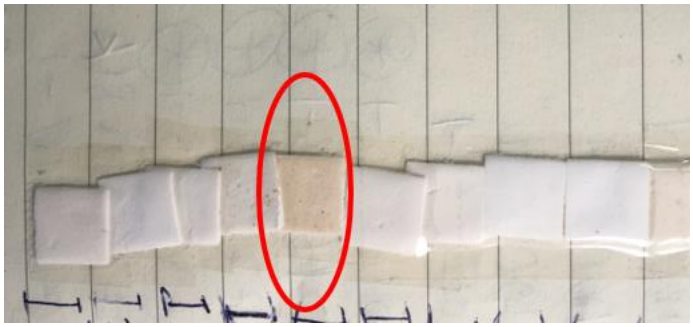

Figure R1. A photo of filters after OC/EC analysis. The red circle highlights an example of abundant metal oxides on the quartz filter after OC/EC analysis, leading to browning of the filter.

6) Section 3.2.1. The reviewer believed the OP mentioned here is reflectance-based. However, Figure 5&6 rely heavily on laser transmittance readings. It is suggested to use OP_R instead of OP to avoid any confusion that may cause.
7) Section 3.2.1 needs more in-depth analysis. The analysis presented here shows that low initial laser T (by Sunset analyzer) is associated with OP_R=0. There could be two possibilities leading to low initial laser T. One possibility (case 1) is a result of high native EC on the filter. That was likely a result of the high surface loading by the IMPROVE sampler, which was mainly designed for remote locations with low PM concentrations. The IMPROVE sampler (22.8 L/min) has a higher flow rate comparing to the SASS sampler (6.7 L/min), which was previously used by CSN before the sampler transition. Besides, the deposit filter area of the IMPROVE sampler (3.53 cm^2) was smaller than the SASS sampler (11.3 cm^2). As a result, the surface loading of the IMPROVE sampler was higher than SASS sampler by a factor of 11, leading to a higher occurrence of low initial laser T. The situation was worsened for CSN samples, which were coming from urban sites with high PM loadings. That poses a challenge for the optical correction in TOA. These should be mentioned and discussed. The second possibility (case 2) is due to high light-absorbing metal oxides on the filter. It is worth investigating the contribution of different causes. For example, how many percentages from case 1, case 2, and both cases 1&2? It would be interesting to see how often that case 1 and case 2 occurred at the same time.
8) Following the last question, as shown in Figure 6d, for the low initial laser T samples (say initial laser T by Sunset analyzer<1000), it seems that half of the samples end up with OP_R=0, but the other half end up with OP_R>0. Why these low initial laser T samples end up with different OP_R? Considering that initial laser T was already low, further darkening of the filter due to charring was likely out of the dynamic range of the optical system, why some samples still get OP_R>0?
9) The authors may consider providing a recommended threshold on the useable range of the laser reading (could be on an absolute or a relative scale). That would be a useful indication for the possible saturation of the optical system due to high loading samples.
10) Section 3.2.2. The authors mentioned that the duration of each step in the IMPROVE_A protocol was concentration-driven. It would be interesting to examine the distribution of duration difference of each carbon fraction between Sunset and DRI instruments. That could be one of the possible sources of carbon fraction divergence between the two instruments.

Technical comments:

1) Line 41. Please cite the latest version of the IPCC report (AR5).
2) Figure 2. Please specify the sample numbers of each plot. Please also add DRI2001 and DRI2015 annotations directly on each plot for easy reference.
3) Figure 6 d&e. Try normalized histograms. The sample size is quite different for the three groups. It is difficult to see the distribution, especially for Figure 6e.