

Reviewer 1's comments

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.

We thank the reviewer for his/her comments on our paper. In the following we provide our point-to-point responses, as well as list our revisions to the text to address these comments.

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.

The change in the carbon analyzer in 2018 is a result of a change in the analytical laboratory that does CSN carbon analysis. In the revised manuscript (Line 63-69) we have added the following text to briefly explain the instrument transition:

“As shown in Figure 1, in the beginning of 2016, TOA carbon analysis for CSN transitioned from using the Desert Research Institute (DRI) Model 2001 analyzers (termed “DRI-2001” hereinafter) to DRI Model 2015 multi-wavelength analyzers (termed “DRI-2015” hereinafter) as a result of instrument upgrade, and again in October 2018, CSN TOA transitioned from using DRI-2015 analyzers to Sunset Laboratory Model 5L analyzers (termed “Sunset” hereinafter) due to change in the analytical laboratory (from DRI to UC Davis).”

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.

We thank the reviewer for the positive feedback.

3) Line 180. The author stated that the mass loading of DRI2001 data was back-calculated from the air concentration data using a constant sample volume (33 m³). 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 DRI2001 data.

We agree that sample-specific volume data, if exists, would be superior to a constant volume to use in calculating mass loadings from concentrations. Nevertheless, such data was unavailable and therefore a constant sample volume was used instead. The uncertainty as a result of this is small given the stringency of CSN operational tolerances for flow rate and sample duration. In CSN, sample is invalidated if the sample flow rate is outside 10% of the nominal flow rate and/or if the sample time is more than 25 hours or less than 23 hours.

We have added a statement in the revised text (Line 197-199) to acknowledge this uncertainty: *“The use of nominal instead of the actual sample volume adds little uncertainty, given the stringency of CSN operational tolerances for flow rate and sample duration.”*

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.

We agree that such a comparison would be useful as it is true that the IMPROVE_A protocol allows for some play in details such as temperature ramping rates and criteria for advancing to the next stage. Unlike most TOA protocols, IMPROVE_A does not pre-specify the temperature profile as the same fixed function of time for each individual sample, but instead leaves it contingent on individual sample compositions and loadings. As Chow et al. (2007) explain, "Temperature is ramped to the next step when the FID [or NDIR] response returns to baseline or remains constant for more than 30 sec; the residence time at each plateau is longer for more heavily loaded samples." Unremarked differences in implicit tolerances for ramping rates, and for determining "return to baseline or remains constant", undoubtedly do contribute some of the differences we observe in different models' reported results. In this work, the time profiles of temperature and evolved carbon for individual samples are available with the Sunset instruments at UCD, but not available from the DRI analyzer. We do agree that a targeted study of such difference between Sunset and DRI analyzers in the future will further refine our understanding of its role in the differences in the analysis results.

We have revised the second paragraph in Section 3.2.2 (Line 449-458) to read "*In addition, details in instrument configuration and operating parameters set by the analysis control program, often invisible and unalterable to end users, can be distinct among TOA models from different manufacturers. As Chow et al. (2007) explain, "Temperature is ramped to the next step when the FID [or NDIR] response returns to baseline or remains constant for more than 30 sec; the residence time at each plateau is longer for more heavily loaded samples." Unremarked differences in implicit tolerances for temperature ramping rates, and for determining "return to baseline or remains constant", undoubtedly contribute some of the differences we observe in different models' reported results. Unfortunately, the time profiles of temperature and evolved carbon for individual samples are not routinely reported by DRI and were not available to us for systematic comparison with those from the Sunset instruments at UCD.*"

We have also added a statement in the revised text (Line 483-486) as an implication of this study that reads:

"It should also be noted that the IMPROVE_A protocol allows for some play in details such as temperature ramping rates and criteria for advancing to the next stage. A targeted study of such difference between Sunset and DRI analyzers in the future will further refine the understanding of its role in the differences in the analysis results."

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.

Thanks for the comment. We agree and are aware that the presence of metal oxide on filter samples is one of the possible causes of low final transmittance/reflectance. During our routine operation of carbon analysis, samples that show orange/red color after TOR analysis due to high Fe abundance are flagged

‘ME-1’ (or ‘m2’ in DRI flagging system). Similarly, samples that show gray/black color after analysis due to residual EC are flagged ‘ME-2’ (or ‘m5’ in DRI flagging system). We have confirmed that the vast majority of the low final T/R samples have the “ME-2” flag.

We have added a statement in the revised text (Line 397-398) that reads:

“Filters with this type of optical profile are black in color before analysis and remain gray/black after analysis.”

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.

We thank the reviewer for the suggestion. In order to avoid any confusions, we have revised Figure 6 (and the new Figure 7) to show reflectance response instead of transmittance. With this change, we keep OP as is, and have made it clear that the optical correction method is by reflectance by replacing “optical charring correction” with “reflectance charring correction” in places where confusion might occur.

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 cm²) was smaller than the SASS sampler (11.3 cm²). 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.

Thanks for the comment. We have added a statement to the revised text (Line 418-421) to mention the first possibility raised by the reviewer that reads: *“The high occurrence of samples with OP=0 in CSN likely results from high sampled air volume, small filter surface area, and the closeness of sampling sites to emission sources, leading to concentrated strong absorbing materials (i.e., EC) on filter samples and posing a challenge for TOA analysis.”*. The second possibility can be ruled out, as discussed in our response to comment 5 above.

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?

This figure is showing data from five Sunset instruments with varying laser sensitivity and intensity (also see response to the next comment). It is possible for some instrument with higher laser sensitivity and a lower baseline to detect some OP even with low initial laser T signal.

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.

We agree that it would be useful to develop a lower laser response threshold. However, the laser response is not standardized across the analyzers, as the actual laser reading (in arbitrary unit) is dependent on not only the optical properties of the sample but also the laser intensity, the geometry of the laser/oven/detector setup, and the response of the photodiode, making it difficult to cross-compare laser response among analyzers of the same model or among different models. As an alternative, in the EPA Air Quality System (AQS) database, the CSN samples with saturated laser problems (samples with OP = 0) are attached a qualifier code (“LJ”) to warn the end users that the reported OC and EC values have larger uncertainties due to laser saturation.

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.

Please see our response to comment 4.

Technical comments:

1) Line 41. Please cite the latest version of the IPCC report (AR5).

Citation has been updated.

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.

Sample numbers have been added to the captions of both Figure 2 and 3 (please note that Figure 2 doesn't include data from DRI-2001). The legend of Figure 3 (which is the DRI model annotation) has been moved to the top of the figure for easier referencing.

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.

Figure 6b and 6c are essentially the normalized distribution. In view of the comment, we have removed Figure 6d and Figure 6e which do not add additional information to Figure 6b and Figure 6c (new Figure 7a and 7b).