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

Review:

This paper discussed the OCEC measurements inter-comparison results obtained from the Chemical Speciation Network (CSN). There are two sets of inter-comparison, but in general, they are comparison between OCEC results obtained from a Sunset instrument vs. DRI analyzer. The main difference between the two sets of measurements lies in the technical details, which are the different analyzer model and the date coverage of the measurements. This is supposed to be a useful inter-comparison exercise and will help the network to understand the impact when switching instrument and how the long-term measurements can be interpret when using it for trend analysis. This work certainly is valuable, however, I find it needs a major revision to re-organize the paper better before it can be published. This is rather confusing and I don’t think there is enough interpretation regarding the actual comparison between the two sets of measurements in a reader perspective. There are many aspects the authors are presenting in this paper, but they are not presented in a way that is easy to follow. I have summarized a list of things that can help improve this manuscript before it is published in AMT.

General comment:

In this paper, the authors rely a lot of statistics and try to interpret the difference (or bias) between the two data sets looking at all data as a population. It is one way of looking at things and to get a sense of the magnitude of this bias in all measurements as a whole. But for ambient measurements, sometimes it is also important to understand the time series of this deviation and to really understand the reason behind the bias. For example, does the bias is higher in summer than winter? Or is always consistent over the course of the sampling period? During the analysis, the authors discovered a significant number of samples that have no observable or detectable “OP”. More investigation is needed to understand if this may impact the data differently for example over different seasons.

In order to obtain the correct OC and EC, charred OC has to be taken into account. The Sunset instrument laser detector detects the transmittance whereas the DRI analyzer has the option to use either transmittance or reflectance. I could not find where the authors discuss about whether they are using transmittance or reflectance to determine charred OC (or OP).

Also, in multiple places when OC and EC were discussed, are they both corrected for charred OC? Any comparison for OC and EC between the two data sets without accounting for charred OC is meaningless.

The authors introduce a new term, referred to as “SRD” or scaled relative difference. I understand this may have some statistical value, but this is not a typical common term that all readers will connect to when they read. This is particularly an issue when readers looking at various graphs generated from this value and try to understand or interpret its meaning. One way to improve this (if the authors insist of keeping this analysis) is to do a better job in explaining this term other than by defining it again. In other words, how readers should interpret this term for its magnitude. For example, when one express absolute difference between two numbers, readers understand the value of 0 means there is no bias. Or large positive or negative values means certain set is higher or lower. In comparison, SRD (because of the way it is defined), it does not easily translate.

Also, is this SRD term mostly used for interpreting the comparison between measurements with replicates? It is not clear from the heading of the sub sections to me.

Specific comment:

p.7, line 275-285. It is observed that Sunset TC is generally agree with DRI TC, however, Sunset OC was found lower than DRI OC (and the opposite is true for EC). This is obviously due to the difference in OCEC split but is could also due to the fact that one instrument used transmittance to determine charred OC whereas the other one use reflectance. I couldn't find where this is discussed or whether the authors are using data to ensure the consistency (i.e., all corrected data are based on transmittance or reflectance).

p.9, line 339. Determining the correct OC and EC really requires the correction of OP. So what is the purpose of showing the "uncorrected" OC and EC when you know OP has to be taken into consideration?

Also, by definition, OP is subtracted from EC and added to OC to determine the final EC and OC. Therefore, in comparing the "uncorrected" and "corrected" OC, one should expect OC and EC be shifted in the same magnitude but in opposite direction. How come this does not seem to be the case in Fig. 4? Is this because it is plotted as a "relative" difference rather than absolute difference? Would absolute difference be more meaningful than the relative difference in this case?

The data in Fig. 4 are sorted according to the "mass loading percentage". How is this defined? Do the authors mean the amount of OC or EC relative to TC? So what is the advantage of plotting the data this way rather than sorting them by the absolute EC and OC mass?

p.9, line 348-350. The authors suggested for EC measurements with large SRD were samples with no instrumentally detected OP. What is considered "large"? Better to give a range or value or threshold. Does this represent bars with high "relative difference" in Fig 4?

In addition, the authors said there are consider number of samples with no detectable OP, if I understand this correctly, no "OP" means the "corrected" OC (or EC) and "uncorrected" OC (or EC) are equal, correct? However, I don't ever see any evidence that there are any samples with "non-detectable OP" in Fig. 4. Fig. 4 shows that there are always a considerable amount of OP in all samples!

p.9, line 355-359. The authors suggested that the EC bias almost solely originated from the thermal effect and I don't understand the argument supporting this statement. In Fig. 5, I only see the sub group of measurements with "no OP" (i.e., no thermal correction) has even higher relative difference than the group with considerable amount of OP. Or may be I don't understand how to interpret this graph. Plotting in relative difference could be one issue. If the authors' statement is true, I would expect for the sub-group of measurements with no OP should give you no bias between the two data set. Is this the case?

p.9, line 360-386. This whole paragraph may be should not belong in main text. Probably is better to be included in supplementary information and then just refer to it in the main text. This paragraph is mostly to explain Fig. 6a, however, I would suggest the authors also include the detector signal, so the authors would better understand how the OP peak compared to OC or EC and aid the interpretation of the thermograms. Only the laser response in Fig. 6 does not help much.

Fig. 6b to e. What is the purpose of these? How does the different shapes in Fig. b and c are supposed to mean and what we should expect? I really don't think these graphs really aid much in terms of understanding what is going on regarding the "OP = 0" issue. The authors can however keep them in supplementary info if they want. What would really help is to actually include a completed thermograms including both laser response and detector response. Only that will allow the readers to understand how the samples evolve over the course of the analysis in a typical situation when OP=0 and OP>0 and may also include the case for blank as reference.