

Response to review #2

Generally Comments

Typically the EC tracer method, when used in estimating the secondary organic carbon (SOC), relies on three conditions– 1) the relatively constant $(OC/EC)_{pri}$ over the period of study; 2) the random nature of SOC formation relative to EC; and 3) a subset of dataset without significant SOC contributions. The $OC/EC_{10\%}$ or OC/EC_{min} essentially utilize the subset in Condition #3 to derive the $(OC/EC)_{pri}$ if it does have a unique value. Any deviations from the conditions as well as measurement uncertainties will lead to bias in determining $(OC/EC)_{pri}$. In some environments where SOC dominates, the third condition is generally impossible to be met. This study, through an extensive test, shows that the third condition is not necessary in calculating $(OC/EC)_{pri}$, if an algorithm, i.e., minimum R^2 (MRS), is used looking for $(OC/EC)_{pri}$ that yields SOC least correlated with EC. Without further examinations, the reviewer thinks that MRS is probably mathematically rigorous for any datasets satisfying the first two conditions and, additionally, with sufficient size and accuracy. It can perform better than $OC/EC_{10\%}$ or OC/EC_{min} most of the time because Condition 3 is fortuitous, as described by the authors.

While the reviewer agrees that MRS should be used instead of $OC/EC_{10\%}$ or OC/EC_{min} in calculating SOC, particularly for a large dataset which can support meaningful correlation analysis, MRS does not solve fundamental problems in the EC tracer method. The $(OC/EC)_{pri}$ is by no means constant, as it varies with source contributions from day to day and season to season. SOC is likely correlated with EC because in urban areas many SOC precursors originate from the same combustion sources as EC. This paper demonstrates that when Conditions 1 and 2 are in doubt, MRS produces erroneous results. MRS results are also sensitive to measurement uncertainty that impacts the correlation coefficients. These limitations, however, are not emphasized adequately in the abstract, which sounds almost like MRS has tackled all these issues. These issues, still, can only be solved by using multivariate or chemical mass balance analysis with additional markers.

Author's Response: Thanks for the very insightful comments. We agree that $(OC/EC)_{pri}$ varied from day to day and season to season in reality and this limitation is intrinsic in the EC tracer method regardless different approaches in implementing the EC tracer method, unless it is applied in a time frame small enough that variations of $(OC/EC)_{pri}$ is almost negligible. Limits posed by the nature of ambient ECOC data are inherent to the EC tracer method and common to all the variants of the EC tracer method. This study focuses on evaluating different $(OC/EC)_{pri}$ determination approaches within the EC tracer method, with the aim to identify the best approach in applying the EC tracer method. We have revised the wording in the abstract and in main text to emphasize the limitations of the EC tracer method and the MRS approach. Please see below the specific revisions in our point-by-point response to reviewers' comments.

Specific Comments

Abstract: Please describe the assumptions of MRS, datasets that are suitable for MRS analysis, and potential errors while in the same time shortening the abstract. Just saying MRS is better than $OC/EC_{10\%}$ or OC/EC_{min} is not meaningful because all the three could be very wrong in some cases.

Author's Response: We have made the following revisions in the abstract to clearly state the assumptions of MRS.

“The hypothetical $(OC/EC)_{pri}$ that generates the minimum $R^2(SOC, EC)$ then represents the actual $(OC/EC)_{pri}$ ratio if variations of EC and SOC are independent and $(OC/EC)_{pri}$ is relatively constant in the study period.”

“...MRS provides an unbiased SOC estimation when measurement uncertainty is small. MRS results are sensitive to the magnitude of measurement uncertainty but the bias would not exceed 23% if the uncertainty is controlled within 20%.”

We also shortened slightly the abstract by condensing a few sentences and removing the following sentence (this background information is spelled out in the introduction section).

~~“The general concept embodied in the MRS method was initially proposed by Miller et al (2005), but has not been evaluated for accuracy or utility since its debut.”~~

Line 97-102: While using simulated data is insightful, it offers no proof. The authors may explore if there is a true “proof” from mathematical or statistical derivations that MRS will yield true $(OC/EC)_{pri}$ if SOC is indeed random and the dataset is big enough. This may also answer the question- how big is big? MRS does not seem suitable for a dataset with only dozens of points.

Author’s Response:

We agree that the simulated data alone does not offer proof, as there is no guarantee that the simulated data capture all the essential features of real-world data. In response to this comment, we conducted a series of sensitivity tests to evaluate the SOC estimation dependency on sample size, which was varied from 20 to 8000. For each sample size, 500 repeat runs were tested, assuming a single value OC/EC_{pri} with a measurement uncertainty of 10%. The results are in Fig. R2, showing the average and the standard deviation for each sample size. The standard variation of SOC bias by MRS decreases with increased sample size while the mean of SOC bias remains a constant small value (2%). The standard variation of SOC bias is $\sim \pm 30\%$ at the lowest tested sample size ($n = 20$), and decreases to less than 15% at $n = 60$ (the sample size of one-year sampling from an every-six-day sampling program) and to less than 10% at $n = 200$. Other scenarios considering OC/EC_{pri} with a distribution and different f_{SOC} are discussed in SI. Figure R2 will be included in the main text.

A new section (as shown below) is added to the manuscript to address the sample size question.

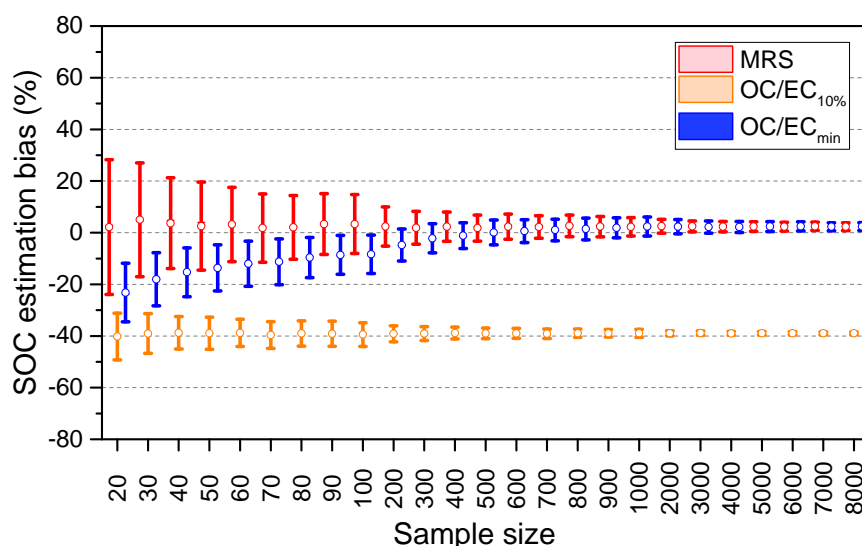


Figure R2. SOC estimation bias as a function of sample size, assuming a fixed relative measurement uncertainty of 10% for OC and EC. For each sample size, 500 repeat runs were conducted. The open circle represents the mean of 500 repeat runs, and the whisker represents one standard deviation. Parameters used for testing: Repeat runs = 500; $N = 8000$; $EC = 8 \pm 4 \mu\text{gC m}^{-3}$; $(OC/EC)_{pri} = 0.5$; $POC = 4 \pm 2 \mu\text{gC m}^{-3}$, $f_{SOC} = 40\%$, and $SOC = 2.67 \pm 1.33 \mu\text{gC m}^{-3}$.

“2.4 Impact of sample size

MRS relies on correlations of input variables and it is expected that MRS performance is sensitive to the sample size of input dataset. This section examines the sensitivity on sample size by the three (OC/EC)_{pri} representations and aims to provide suggestions for an appropriate sample size when applying MRS on ambient OCEC data. Sample sizes ranging from 20 ~ 8000 are tested and for each sample size 500 repeat runs are conducted to obtain statistically significant results. Both Case A (i.e., a constant relative uncertainty of 10%) and Case B (i.e., a constant absolute uncertainty of $\pm 0.2 \mu\text{gC m}^{-3}$ for both OC and EC) are considered. The measurement uncertainties in case B are generated separately by MT following a uniform distribution within the range of $\pm 0.2 \mu\text{gC m}^{-3}$. The measurement uncertainties of POC and SOC are then back-calculated following the uncertainty propagation formula (Harris, 2010) and assuming the ratio of $\varepsilon_{\text{POC}}/\varepsilon_{\text{SOC}}$ is the same as POC/SOC ratio (controlled by f_{SOC}).

The mean SOC bias by MRS is very small (<3%) for all sample sizes while the standard deviation of SOC bias decreases with sample size (Figure 8). The standard deviation of SOC bias is $\sim \pm 30\%$ at the lowest test sample size ($n = 20$), and decreases to less than $\pm 15\%$ at $n = 60$ (the sample size of one-year sampling from an every-six-day sampling program) and to less than $\pm 10\%$ at $n = 200$. Similar patterns are observed between Case A (Figure 8a) and Case B (Figure 8b) for MRS and OC/EC_{10%}. For OC/EC_{min}, a larger bias is observed in Case B than Case A for all sample sizes, as SOC bias by OC/EC_{min} is more sensitive to measurement uncertainty in the range of 0~10% as shown in Figure 7b. The standard deviation of SOC bias by OC/EC_{min} and OC/EC_{10%} both decreases with sample size as shown in Figure 8. The mean SOC bias of OC/EC_{min} decrease with increased sample size while OC/EC_{10%} is insensitive to sample size. The sample size dependency of all three (OC/EC)_{pri} representations is not sensitive to f_{SOC} as shown in Figure S16. Other scenarios considering (OC/EC)_{pri} with a distribution and different f_{SOC} are discussed in SI.”

Line 116-118: How good are the K-S statistics? In other words, how well did the pseudorandom number generator reproduce the statistics in the original dataset?

Author’s Response: The K-S statistics for ambient measured data are shown in Figures S1-S4 (This information is now also mentioned in the main text). In Igor Pro’s Kolmogorov–Smirnov test, D represents the K-S statistic, C represent critical value. If $D < C$, the samples follow the corresponding distribution (e.g., normal or log-normal distribution). The majority of the data can pass the K-S test for log-normal distribution and some exhibit a bimodal distribution. For the performance of the MT pseudorandom number generator, we conduct a series of K-S tests on the generated data for 5000 runs, which show 94.4% data having D small than C (Fig. R3). Hence, we believe the pseudorandom number generator could produce the data following preset characteristics. Figure R3 is added to the SI and referred to in the main text.

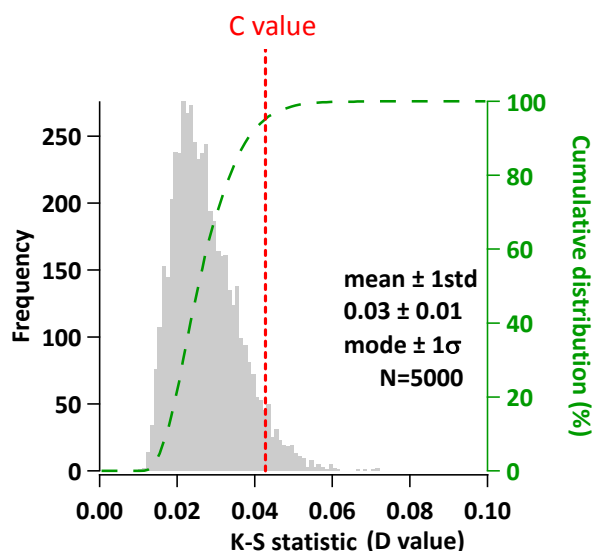


Figure R3 Performance of the MT pseudorandom number generator evaluated by K-S test. The histogram in grey represents D statistic value in K-S test and the red dashed-line represents C. The dash line in green represents cumulative distribution of D. Data with $D < C$, i.e., data that strictly follow the log-normal distribution, account for 94.4% in 5000 runs.

The below text is added to the manuscript to describe whether the pseudorandom number generate reproduce the statistics in the original dataset.

“To verify the log-normality of MT generated data, a series of K-S tests on the generated data for 5000 runs are conducted. As shown in Figure S6, 94.4% of runs pass the K-S test. Hence the performance of MT can satisfy the log-normally distributed data generation requirement in this study”.

Line 126: Eqs. (4)-(5) do not work for all datasets. They are probably asymptotes when datasets are large enough in size.

Author’s Response: We agree that they do not necessarily work for all datasets. The reason for translating mean and standard deviations into μ and σ is that the MT pseudorandom number generator in Igor Pro only accepts μ and σ as input parameters, while mean and standard deviations are the parameters that can be obtained from ambient measurements.

Line 136: Mention here that the case with combustion-related SOC is discussed later.

Author’s Response: Suggestion taken. The text below is included in the revised manuscript:

“The case with combustion-related SOC is briefly discussed in section 3.”

Line 151-152: The results of log-normally distributed $(OE/EC)_{pri}$ should be summarized in the text if possible.

Author’s Response: Suggestion taken. The below text is added to the section 2.2.1:

“Analysis is also performed to evaluate SOC estimation bias as a function of RSD_{EC} , RSD_{SOC} , and f_{SOC} if $(OC/EC)_{pri}$ is considered to be lognormally distributed. Table S2 summarizes the results obtained with adopting most probable ambient conditions (i.e., RSD_{EC} : 50-100%, f_{SOC} : 40-60%). SOC bias by MRS is within 4% when measurement uncertainty is ignored. In comparison, SOC

bias by OC/EC_{\min} is more sensitive to assumption of log-normally distributed $(OC/EC)_{\text{pri}}$ than single value $(OC/EC)_{\text{pri}}$, including the dependency on RSD_{EC} and RSD_{SOC} with varied f_{SOC} .”

Line 220-222: It is not clear if f_{EC1} was varied from sample to sample in a single test or only varied from test to test. If the former, how could you make sure EC1 and EC2 are highly correlated?

Author’s Response: f_{EC1} was varied from test to test. The text is now clarified as below:

“By varying f_{EC1} (proportion of source 1 EC to total EC) from test to test, the effect of different mixing ratios of the two sources can be examined.”

Line 284-286: Since POC and SOC are not directly measured, what is the meaning to simulate their measurement uncertainty?

Author’s Response: Once OC and EC data are considered to have measurement uncertainty, the derived quantities POC and SOC (using Eq (1) and Eq (2)) consequently also have associated uncertainty, which can be calculated following uncertainty propagation principle. For the evaluation of SOC estimation, SOC calculated from the EC tracer method needs to be compared with “true SOC plus associated uncertainty”. That’s the reason why we calculated the uncertainties of POC and SOC

Line 384: How were the six subsets selected?

Author’s Response: With a given one-year data set, there are six possible extractions of daily data sets corresponding to the assumed every-six-day sampling schedule, i.e., set 1: {Day 1, 7, 13,...}, set 2: {Day 2, 8, 14,...}, set 3: {Day 3, 9, 15,...}, etc. The text below is added to clarify this point:

“The one-year data yields six subsets of daily samples, corresponding to six possible schedules of sampling days with the every-six-day sampling frequency. The MRS calculation produces the OC/EC_{pri} in the range of 2.37 – 2.75...”

Line 360-362: Emphasize that this only happens when measurement uncertainties are small.

Author’s Response: Suggestion taken. This sentence is revised as below:

“In the scenarios of a single primary source and two well-correlated primary combustion sources, SOC estimates by MRS are unbiased while OC/EC_{\min} and $OC/EC_{10\%}$ consistently underestimate SOC when measurement uncertainty is neglected. When measurement uncertainty is considered, all three approaches produce biased SOC estimates, with MRS producing the smallest bias. The bias by MRS is less than 25% when measurement uncertainty is within 20% and f_{SOC} is not lower than 20%.”