Supplemental materials for

Determination of primary organic carbon-to-elemental carbon (OC/EC) ratio using ambient OC and EC measurements: Secondary OC-EC correlation minimization method

Cheng Wu[1], Jian Zhen Yu[1, 2, 3]

[1]Division of Environment, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

[2] Atmospheric Research Centre, Fok Ying Tung Graduate School, Hong Kong University of Science and Technology, Nansha, China

[3] Department of Chemistry, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China

Correspondence to: Jian Zhen Yu (jian.yu@ust.hk)

1.   OC and EC Measurements

Semi-continues thermal-optical transmittance (TOT) carbon analyzers (Model RT-4, Sunset Laboratory Inc., Tigard, Oregon, USA) were deployed at three sites in Pearl River Delta (PRD) region, China for hourly $PM_{2.5}$ organic carbon (OC) and elemental carbon (EC) measurement. Briefly, ambient air is first drawn into a sharp-cut $PM_{2.5}$ cyclone (BGI, Waltham, Massachusetts, USA) at the flow rate of 8 Lmin^{-1}, then passing through a parallel plated carbon denuder (Sunset Laboratory, Inc., Tigard, Oregon, USA) to remove gaseous OC, and finally collected onto two back-to-back 16 mm diameter quartz filter (Pallflex, Tissuquartz, 2500-QAT-UP, Putnam, Massachusetts, USA). After 45 min of sampling the particle laden filter is subjected to analysis. The analysis is done in the following 15 min. In thermal-optical analysis, OC is volatized first by programed temperature steps in oxygen-free atmosphere while in the second stage combustion of EC requires presence of oxygen. Evolved OC is converted to carbon dioxide (CO_2) by manganese dioxide catalyst. Determination of CO_2 from oxidized OC and EC is quantified by a non-dispersive infrared (NDIR) detector. During the Helium stage a fraction of OC is pyrolized at high temperature and lead to formation of pyrolized carbon (PC), resulting filter darkened. To minimize this analytical artifact, a tunable red diode laser (660 nm) is used for monitoring the transmittance signal of filter. Formation of PC in He stage decreases the laser signal while combustion of PC and EC in O_2 stage results in recovery of laser signal. When the laser signal in O_2 stage returns to the initial level, the carbon fraction before this point in O_2 stage is quantified as PC, achieving the differentiation of PC and native EC. So all carbon fractions volatized before this point (known as OC/EC split) are regarded as OC (OC evolved in He + PC combusted in O_2) and after as EC.

The three monitoring sites include (1) Nancun, a suburban location in Guangzhou, Guangdong , (2) Tsuen Wan, an urban in Hong Kong, and (3) Mong Kok, an urban roadside in Hong Kong. Figure S1 shows the frequency distributions of OC, EC, and OC/EC ratios at the three sites for the whole year. Figures S2-S4 show the seasonal subsets using the Nancun site as the example. Figure S5 shows frequency distribution of OC/EC ratio during rush hours (7:00-9:00) at the urban roadside location Mong Kok.
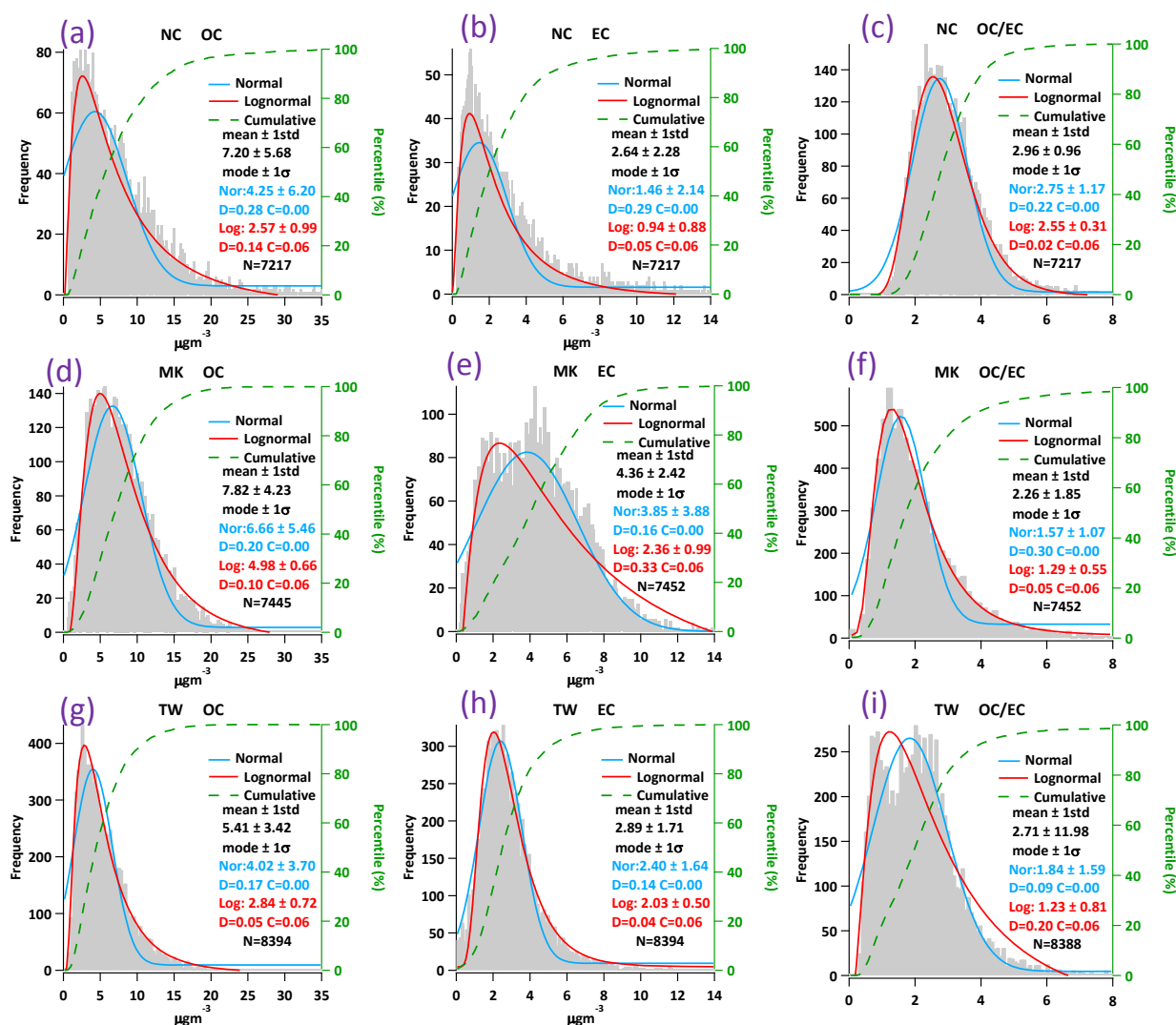
Figure S1. Frequency distributions of OC, EC and OC/EC ratio in ambient samples recorded at three locations in the Pearl River Delta region, China. (a)-(c) Nancun, a suburban site in Guangzhou, (d)-(f) Mong Kok, an urban roadside site in Hong Kong, (g)-(i) Tsuen Wan, a general urban site in Hong Kong. Grey areas represent frequency distribution, green dashed lines represent cumulative frequency distribution, red lines are normal fitting, and blue lines are log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.
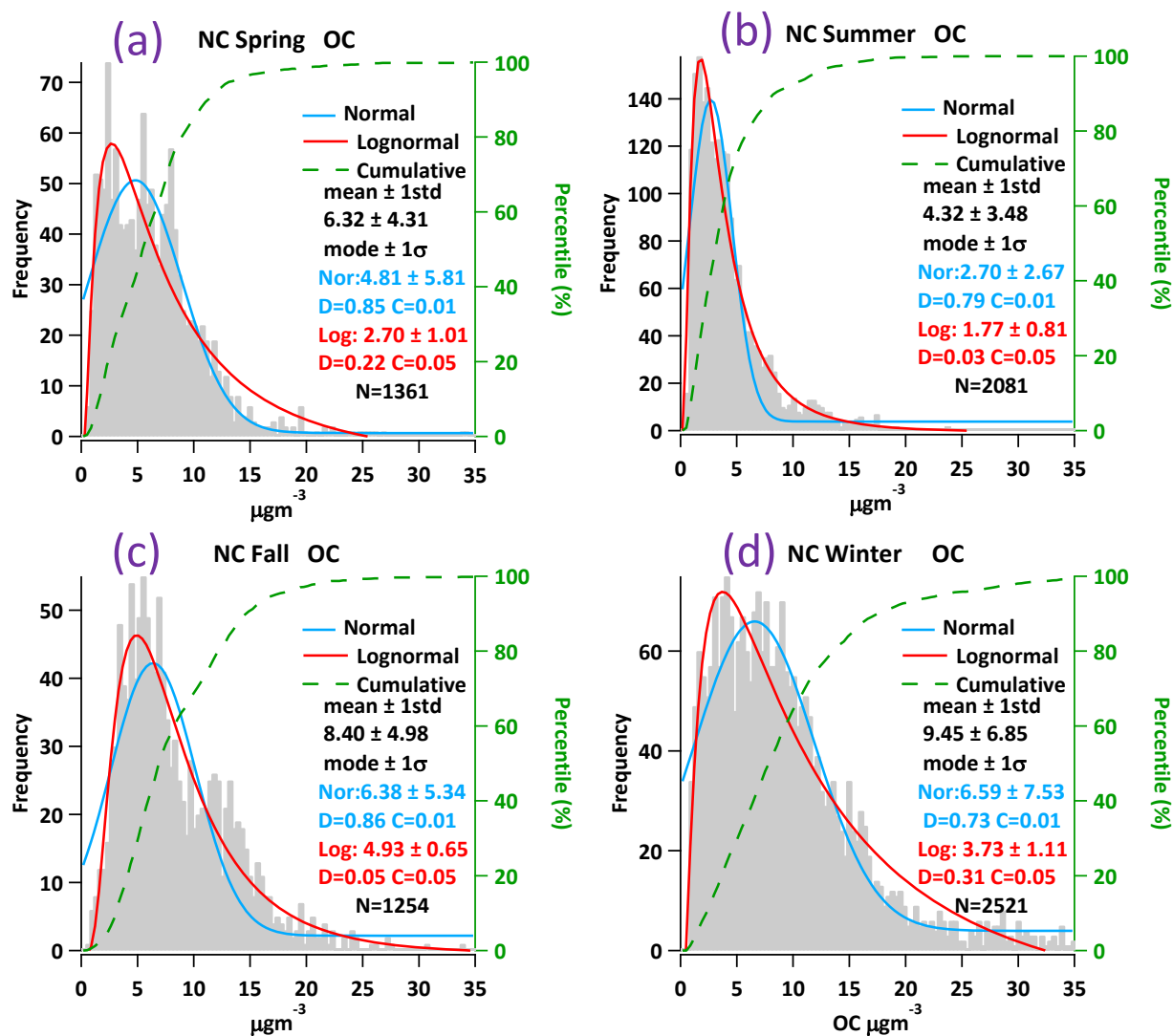
Figure S2. Seasonal frequency distributions of **OC** observed at the suburban site Nancun in the Pearl River Delta region. (a)-(d) represent spring, summer, fall, and winter, respectively. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, red lines represent normal fitting, blue lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.
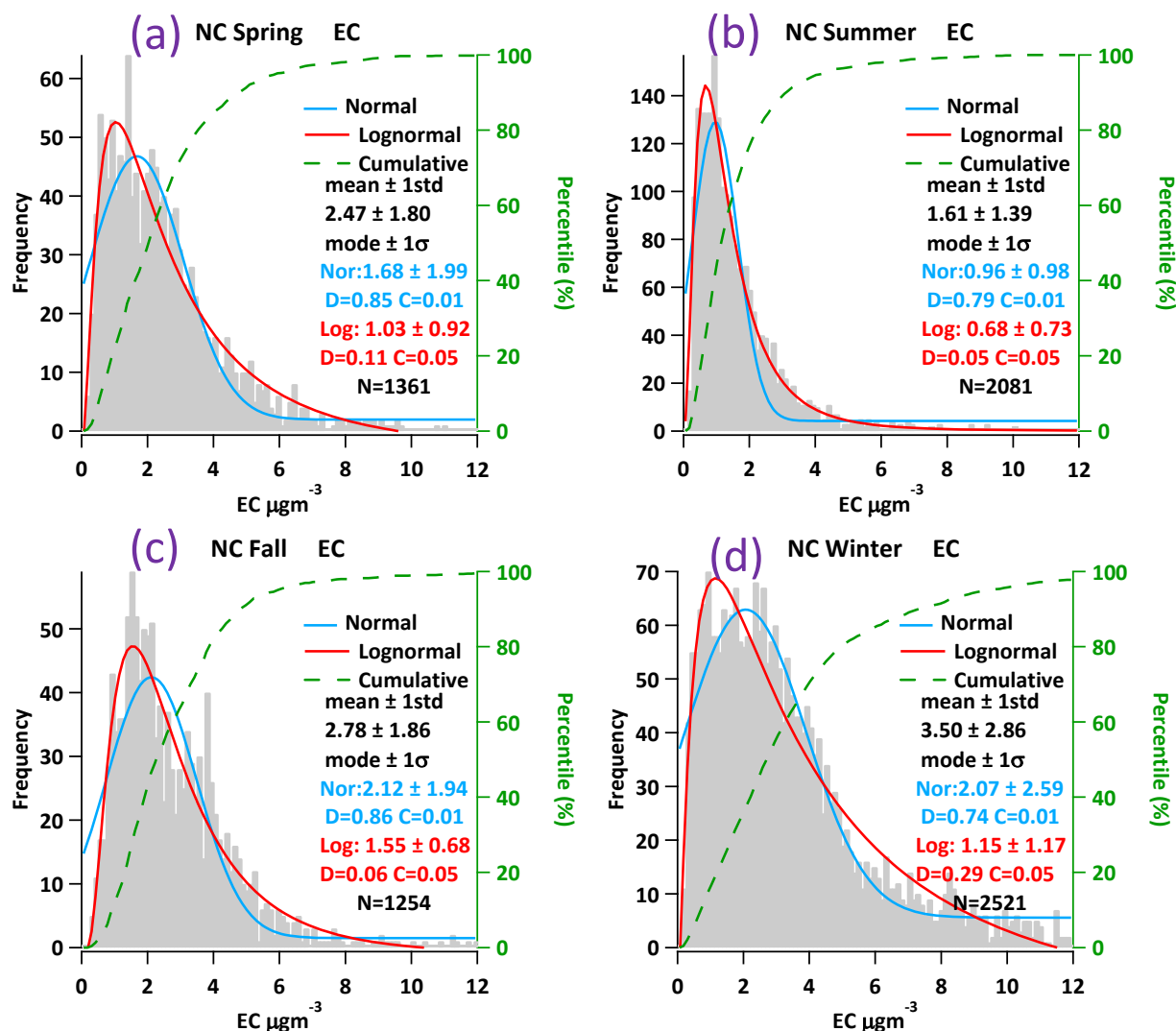
Figure S3. Seasonal frequency distributions of **EC** observed at the suburban site Nancun in the Pearl River Delta region. (a)-(d) represent spring, summer, fall, and winter, respectively. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, red lines represent normal fitting, blue lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.
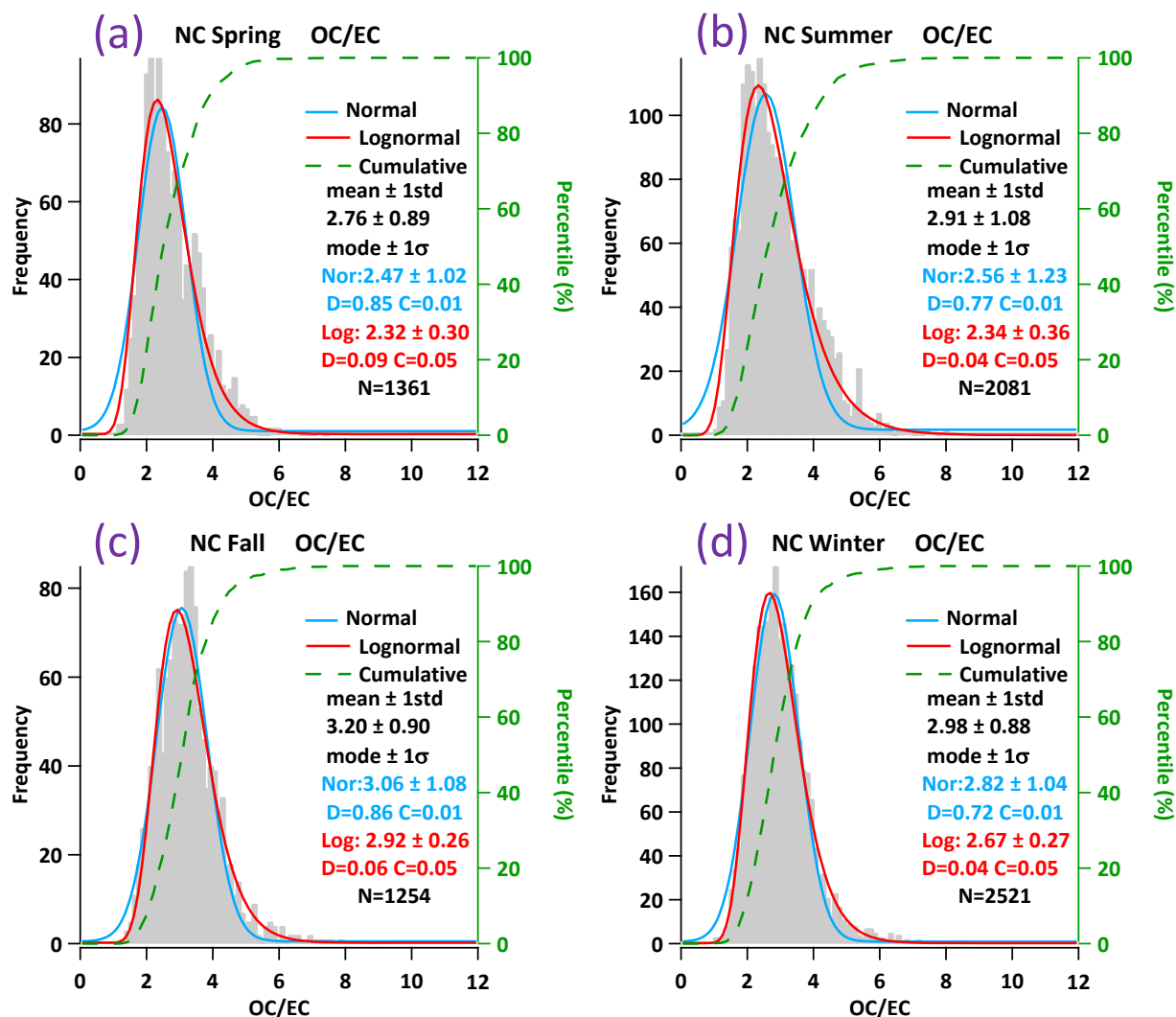
Figure S4. Seasonal frequency distributions of **OC/EC** ratio observed at the suburban site Nancun in the Pearl River Delta region. (a)-(d) represent spring, summer, fall, and winter, respectively. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, red lines represent normal fitting, blue lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.
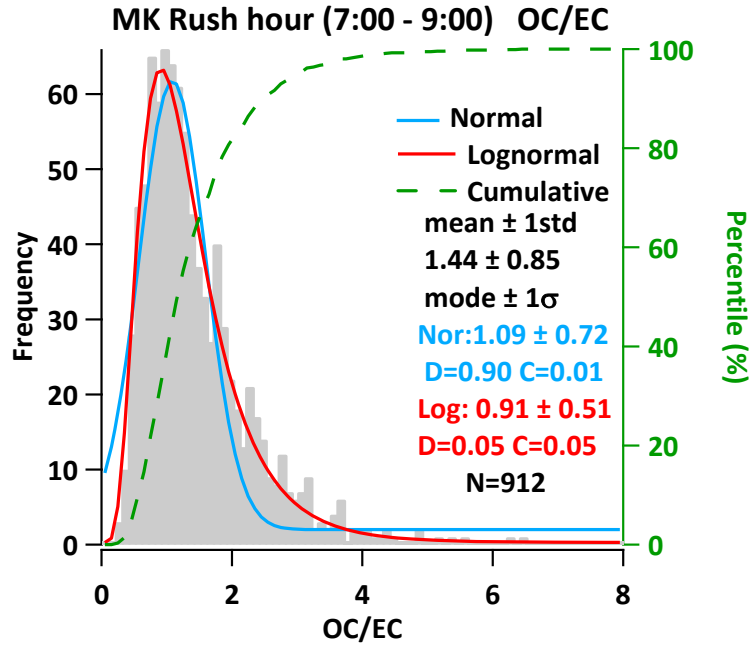
Figure S5. Frequency distribution of OC/EC ratio observed at Mong Kok (MK), a roadside site in Hong Kong. OC/EC ratio of road site MK can represent the (OC/EC)$_{pri}$ in Hong Kong. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, red lines represent normal fitting, blue lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.
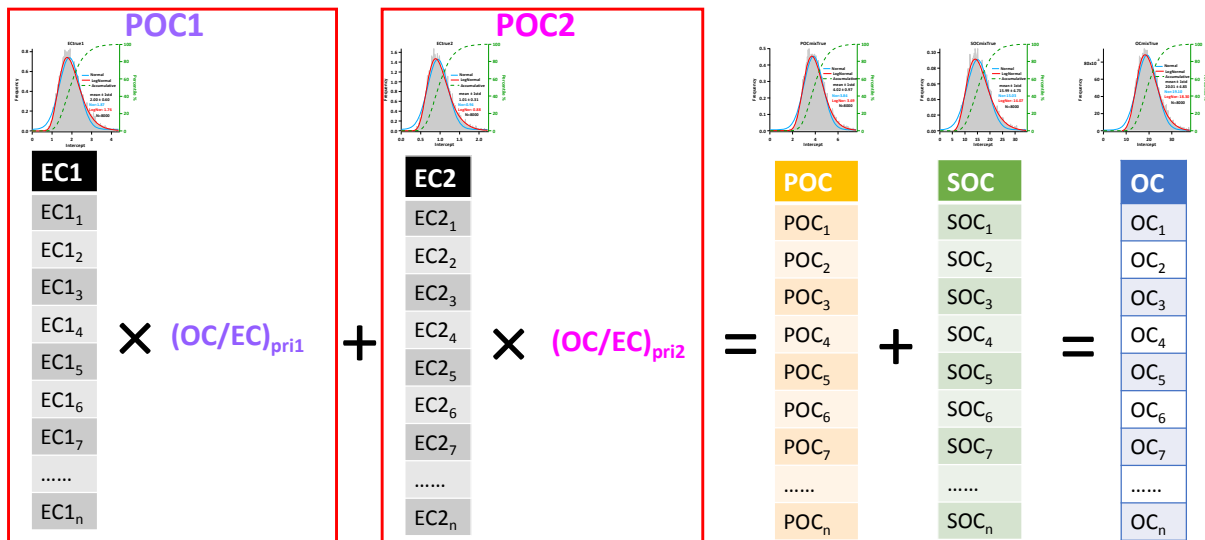
Figure S6. Schematic diagram of pseudorandom number generation for two primary emission sources. The data series (EC$_1$, EC$_2$ and SOC), which are generated by Mersenne twister (MT) pseudorandom number generator, statistically follow log-normal distribution, but the sequence of each data point is randomly assigned.

Single value (OC/EC)$_{pri}$
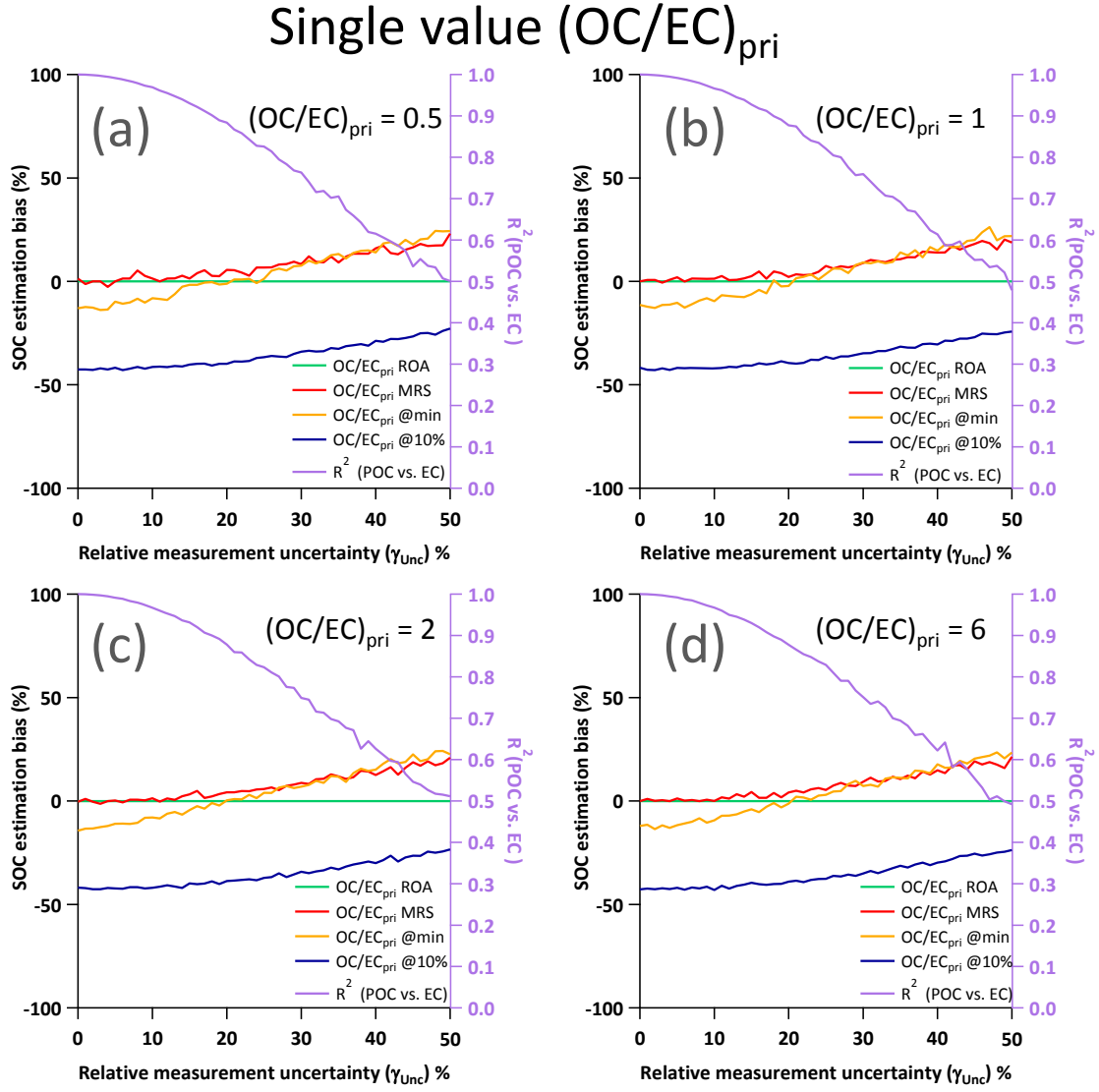
Figure S7. Variation of bias of SOC determination with measurement uncertainty under four different (OC/EC)$_{pri}$ (a) (OC/EC)$_{pri}$ = 0.5, (b) (OC/EC)$_{pri}$ = 1, (c) (OC/EC)$_{pri}$ = 2 , and (d) (OC/EC)$_{pri}$ = 6, demonstrating dependence on γ_{unc} while no effects by (OC/EC)$_{pri}$.

The different representations of (OC/EC)$_{pri}$ include ratio of averages (ROA), minimum R square method (MRS), OC/EC$_{10\%}$, OC/EC$_{min}$. Fixed input parameters: N=8000; EC = 2±1 μgm^{-3}, and f$_{SOC}$=60%.

2. Single source scenario: $(OC/EC)_{pri}$ with lognormal distribution

The procedure of data generation considering lognormal distribution in $(OC/EC)_{pri}$ is illustrated in Figure S8 and implemented by scripts written in Igor Pro. EC is first generated with the following parameters specified: sample size, mean and relative standard deviation (RSD%) of the whole data set . The EC data set statistically follows log-normal distribution, while the sequence of each data point is randomly assigned. Next, a log-normally distributed $(OC/EC)_{pri}$ data set, is generated, with a given sample size, mean and RSD. POC is then calculated by multiplying EC and $(OC/EC)_{pri}$. SOC data is independently generated in a similar way to that for EC. The synthesized OC is then the sum of POC and SOC. OC and EC data sets generated in this way are used to evaluate the accuracy of SOC estimates by different implementations of the EC tracer method.

In ambient atmospheric environments, due to the co-emission of POC and EC, high R^2(POC, EC) (i.e., correlation coefficient between POC and EC) is expected. In order for the simulated data set to retain this relationship, we found that the ratio between the RSD values of $(OC/EC)_{pri}$ and EC data (referred to $\gamma_{\_RSD}$ for convenience) is a key factor and needs to be constrained. Figure S10 shows R^2 (POC, EC) deteriorates as $\gamma_{\_RSD}$ of the simulated data sets increases. At $\gamma_{\_RSD}$ > 20%, R^2 (POC, EC) falls below 0.95. In our numerical experiments, $\gamma_{\_RSD}$ is constrained to be 10% to maintain the high correlation between POC and EC.

We note that the simulated $(OC/EC)_{pri}$ is not a single value, reflecting its variable nature under realistic atmospheric environments. Multiple parameters are available to describe the $(OC/EC)_{pri}$ as summarized in Table S1, including slope obtained through Deming regression (DR), ratio of averages (ROA) POC against EC, median of ratios (MER), mode of ratios (MDR), average of ratios (AOR). It is necessary to select one of these parameters as the reference $(OC/EC)_{pri}$ to evaluate the performance of $(OC/EC)_{pri}$ estimation by MRS. Studies by Chu (2005) and Saylor et al. (2006) both suggest ROA being the best estimator of the expected primary OC/EC ratio. Therefore, the deviations of various $(OC/EC)_{pri}$ representations from ROA as a function of $\gamma_{\_RSD}$ are examined in Figure S9. Increasing bias in MDR, MER and DR are observed as a function of $\gamma_{\_RSD}$ while the deviation of $(OC/EC)_{pri}$ by MRS from ROA is within ±5% throughout a range of $\gamma_{\_RSD}$ from 0 to 100%. This result clearly demonstrates the robustness of MRS in $(OC/EC)_{pri}$ estimation.

As discussed in the conceptual diagram (Figure 3), distribution width and distance between $(OC/EC)_{pri}$ and OC/EC distributions could vary in ambient samples. This is challenging for $OC/EC_{10\%}$ and OC/EC_{min} because they rely on the position of left tail of OC/EC distribution. The accuracy of SOC estimates by $OC/EC_{10\%}$ and OC/EC_{min} methods has a strong dependence on RSD_{EC}, RSD_{SOC} and f_{SOC}, which control the distance and widths of the two peaks. Therefore, sensitivity tests are performed by varying these parameters: RSD_{EC}, RSD_{SOC} and f_{SOC}. As shown in Figure S10, $OC/EC_{10\%}$ and OC/EC_{min} exhibit strong dependence on RSD_{EC} and f_{SOC}. When RSD_{EC} is larger, the SOC underestimation will be smaller. When f_{SOC} <60%, SOC bias can be either overestimation or underestimation as a function of RSD_{EC}. Sensitivity test on RSD_{SOC} and f_{SOC} are examined in Figure S12. When f_{SOC} <20%, the OC/EC_{min} approach overestimates SOC, which increase with RSD_{SOC}. When f_{SOC} >20%, the OC/EC_{min} approach underestimates SOC, which decrease with RSD_{SOC}. The $OC/EC_{10\%}$ approach always underestimates SOC but the degree of underestimation decreases with higher RSD_{SOC}. Sensitivity test on γ_{Unc} and f_{SOC} are examined in Figure S12. MRS and the OC/EC_{min} approach tend to overestimate SOC if γ_{Unc} become higher. SOC underestimation by $OC/EC_{10\%}$ decrease as γ_{Unc} becomes higher. SOC bias values by MRS, $OC/EC_{10\%}$ and OC/EC_{min} are all reduced when f_{SOC} is higher. Considering a reasonable measurement uncertainty (γ_{Unc} < 20%), MRS is shown to provide the best performance, then

followed by OC/EC$_{min}$. OC/EC$_{10\%}$ exhibits the largest bias among three approaches. SOC bias dependency on (OC/EC)$_{pri}$ and EC concentration is also tested as shown in Figures S13 and S14. The results show that the SOC estimate bias by the MRS, OC/EC$_{min}$ and OC/EC$_{10\%}$ methods due to elevated γ_{Unc} does not depends on (OC/EC)$_{pri}$ and EC concentration, except OC/EC$_{min}$ when (OC/EC)$_{pri}$ =6 (Figure S13d). Results of sensitivity tests discussed above are summarized in Table S2 considering most possible ambient conditions.

Table S1. Various ways representing (OC/EC)$_{pri}$ and abbreviation of terms used in this numerical study

Representation of (OC/EC)$_{pri}$	Input parameters	Remarks
average of ratios(AOR)	POC EC	$\overline{POC/EC}$
median of ratios(MER)	POC EC	$\widetilde{POC/EC}$
mode of ratios(MDR)	POC EC	
ratio of averages (ROA)	POC EC	$\overline{POC}/\overline{EC}$
deming regression (DR)	POC EC	Slope of regression
minimum R square method (MRS)	OC EC	MRS
OC/EC$_{min}$	OC EC	Minimum OC/EC
OC/EC$_{10\%}$	OC EC	OC/EC at 10%

Table S2. Summary of numerical study results in scenarios 1 (S1) considering (OC/EC)$_{pri}$ with lognormal distribution under most probable general ambient conditions*. "+" represent SOC overestimation, "-" represent underestimation.

	Tested parameter	SOC bias due to tested parameter		
		MRS	OC/EC$_{min}$	OC/EC$_{10\%}$
	RSD$_{EC}$	±4%	±20%	-45% ~ -40%
Scenario 1	RSD$_{SOC}$	±4%	-20% ~ 0%	-50% ~ -20%
Single source	f$_{SOC}$	±4%	±20%	-50% ~-20%
	γ_{Unc}	+10%	+25%	-45%

*Most probable general ambient conditions: RSD$_{EC}$ 50~100%; f$_{SOC}$ 40~60%; γ_{Unc} 20%;
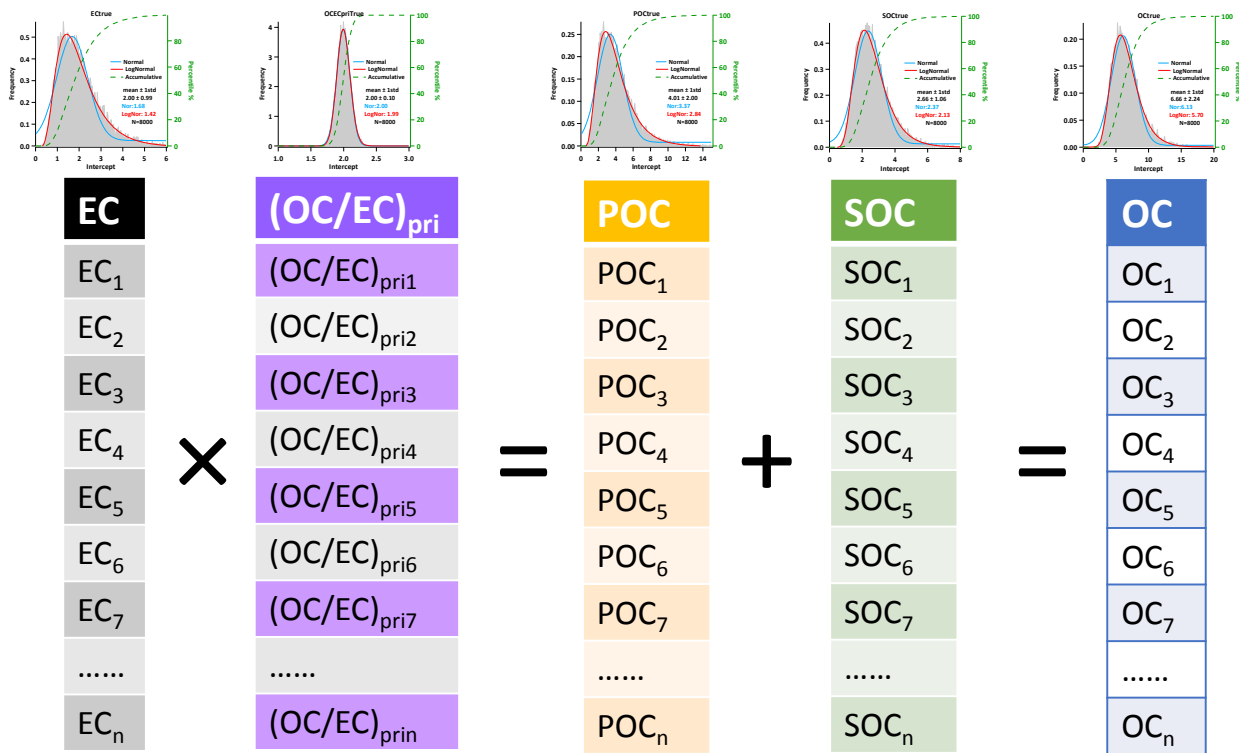
Figure S8. Schematic diagram of pseudorandom number generation. The data series (EC, (OC/EC)$_{pri}$ and SOC), which are generated by Mersenne twister (MT) pseudorandom number generator, statistically follow log-normal distribution, but the sequence of each data point is randomly assigned.
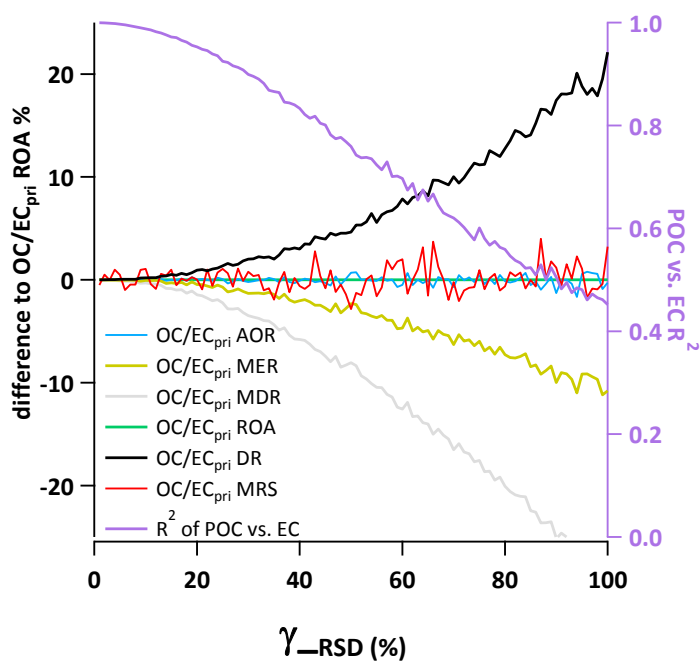
Figure S9. Deviation of various representation of (OC/EC)$_{pri}$ from the ratio of averages (ROA) as a function of (OC/EC)$_{pri}$ RSD to EC RSD ratio (γ_{RSD}). Representation of (OC/EC)$_{pri}$ including: average of ratios(AOR), median of ratios(MER), mode of ratios(MDR), deming regression (DR), minimum R square method (MRS). Input configurations: N=8000 EC = 4±2 µgm^{-3} (OC/EC)$_{pri}$ = 2 SOC/OC ratio (f$_{SOC}$)=40% SOC = 5.3±2.6 µgm^{-3}

(OC/EC)$_{pri}$ with lognormal distribution
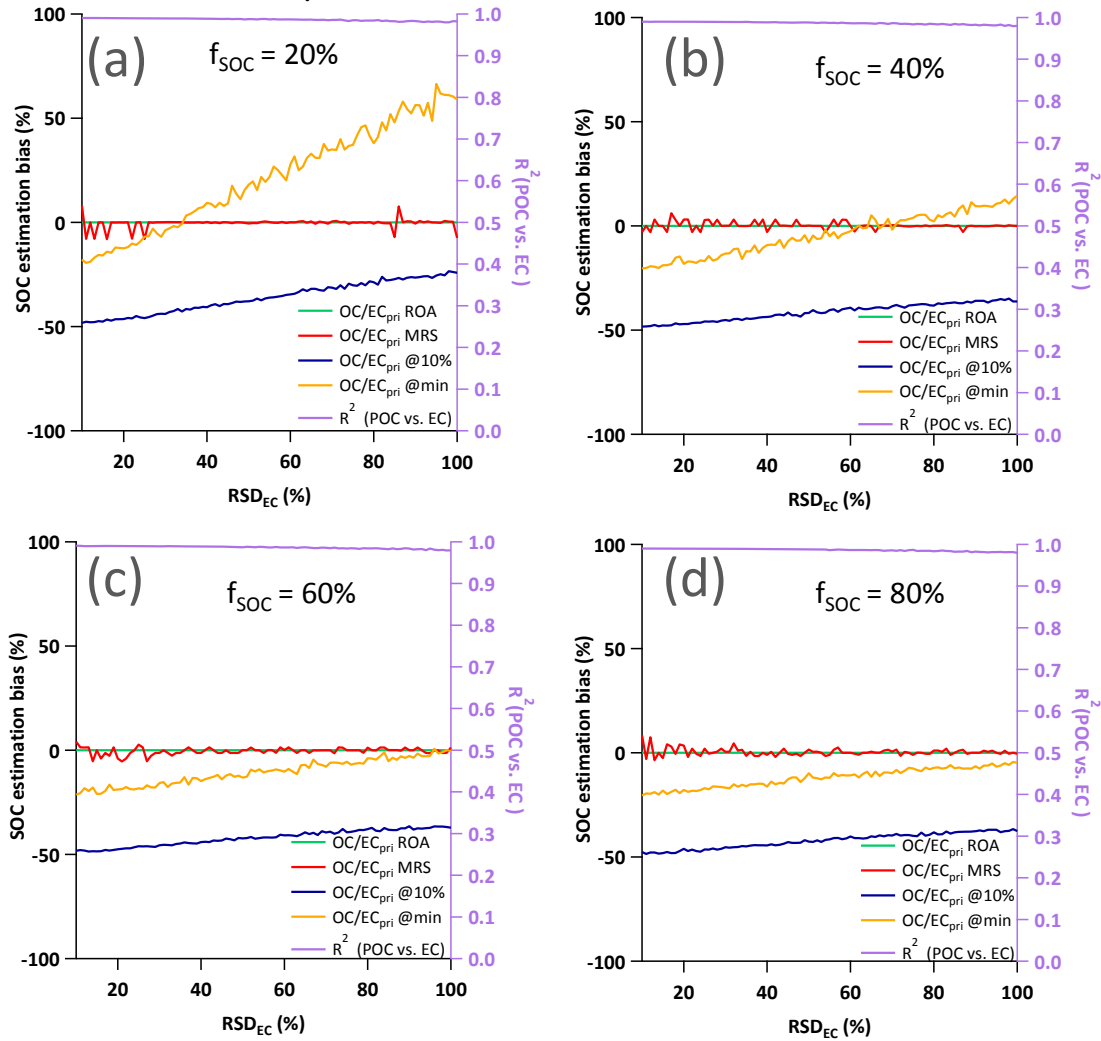
Figure S10. Bias of SOC determination as a function of RSD$_{EC}$ and f$_{SOC}$ by different representation of (OC/EC)$_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), OC/EC$_{10\%}$, OC/EC$_{min}$. Fixed input parameters: N=8000; EC = 2±1 µgm^{-3}, and (OC/EC)$_{pri}$=0.5±0.025. f$_{SOC}$ is varied, with (a) f$_{SOC}$ =20% SOC = 0.25±0.13 µgm^{-3} (b) f$_{SOC}$ =40% SOC = 0.67±0.33 µgm^{-3} (c) f$_{SOC}$ =60% SOC = 1.5±0. 75 µgm^{-3} (d) f$_{SOC}$ =80% SOC = 4±2 µgm^{-3}

$(OC/EC)_{pri}$ with lognormal distribution
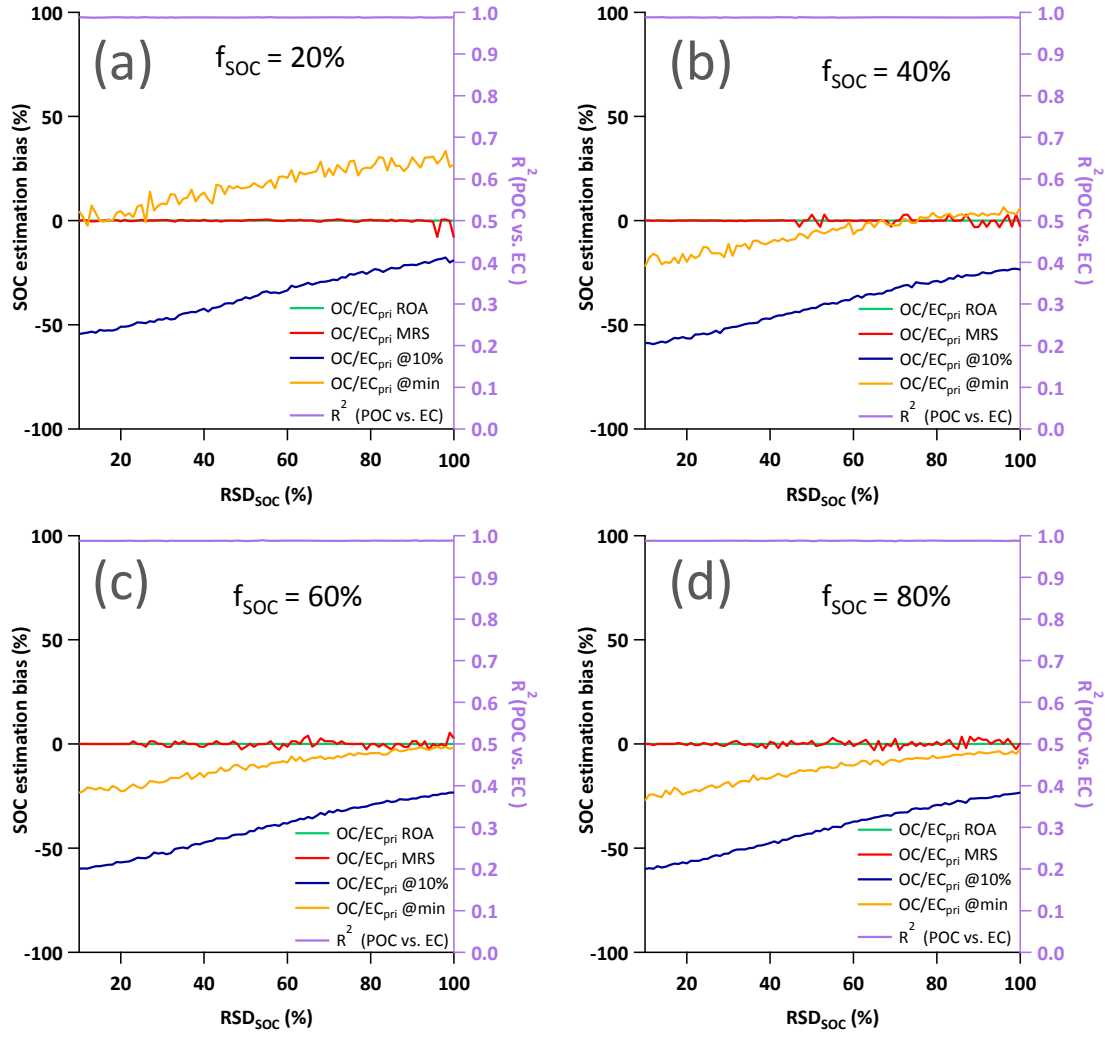
Figure S11. Bias of SOC determination as a function of RSD_{SOC} and f_{SOC} by different representation of $(OC/EC)_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), $OC/EC_{10\%}$, OC/EC_{min}. Fixed input parameters: N=8000; EC = 2±1 μgm^{-3}, and $(OC/EC)_{pri}$=0.5±0.025. f_{SOC} is varied, with (a) f_{SOC} =20% SOC = 0.25±0.13 μgm^{-3} (b) f_{SOC} =40% SOC = 0.67±0.33 μgm^{-3} (c) f_{SOC} =60% SOC = 1.5±0. 75 μgm^{-3} (d) f_{SOC} =80% SOC = 4±2 μgm^{-3}

(OC/EC)$_{pri}$ with lognormal distribution

Figure S12. Bias of SOC determination as a function of relative measurement uncertainty (γ_{Unc}) and (OC/EC)$_{pri}$ by different representation of (OC/EC)$_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), OC/EC$_{10\%}$, OC/EC$_{min}$. Fixed input parameters: N=8000; EC = 2±1 µgm^{-3}, and (OC/EC)$_{pri}$=0.5±0.025. f$_{SOC}$ is varied, with (a) f$_{SOC}$ =20% SOC = 0.25±0.13 µgm^{-3} (b) f$_{SOC}$ =40% SOC = 0.67±0.33 µgm^{-3} (c) f$_{SOC}$ =60% SOC = 1.5±0. 75 µgm^{-3} (d) f$_{SOC}$ =80% SOC = 4±2 µgm^{-3}

(OC/EC)$_{pri}$ with lognormal distribution
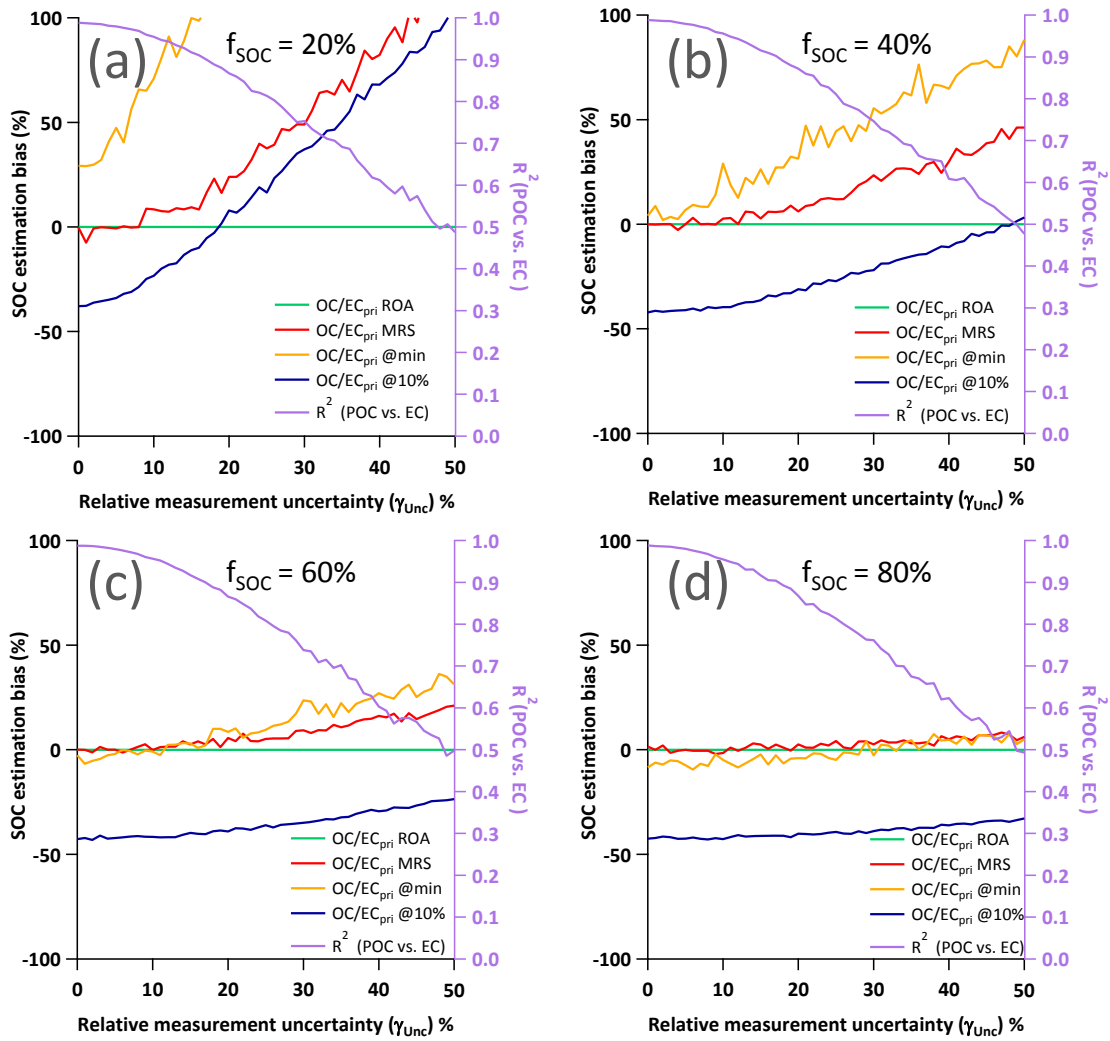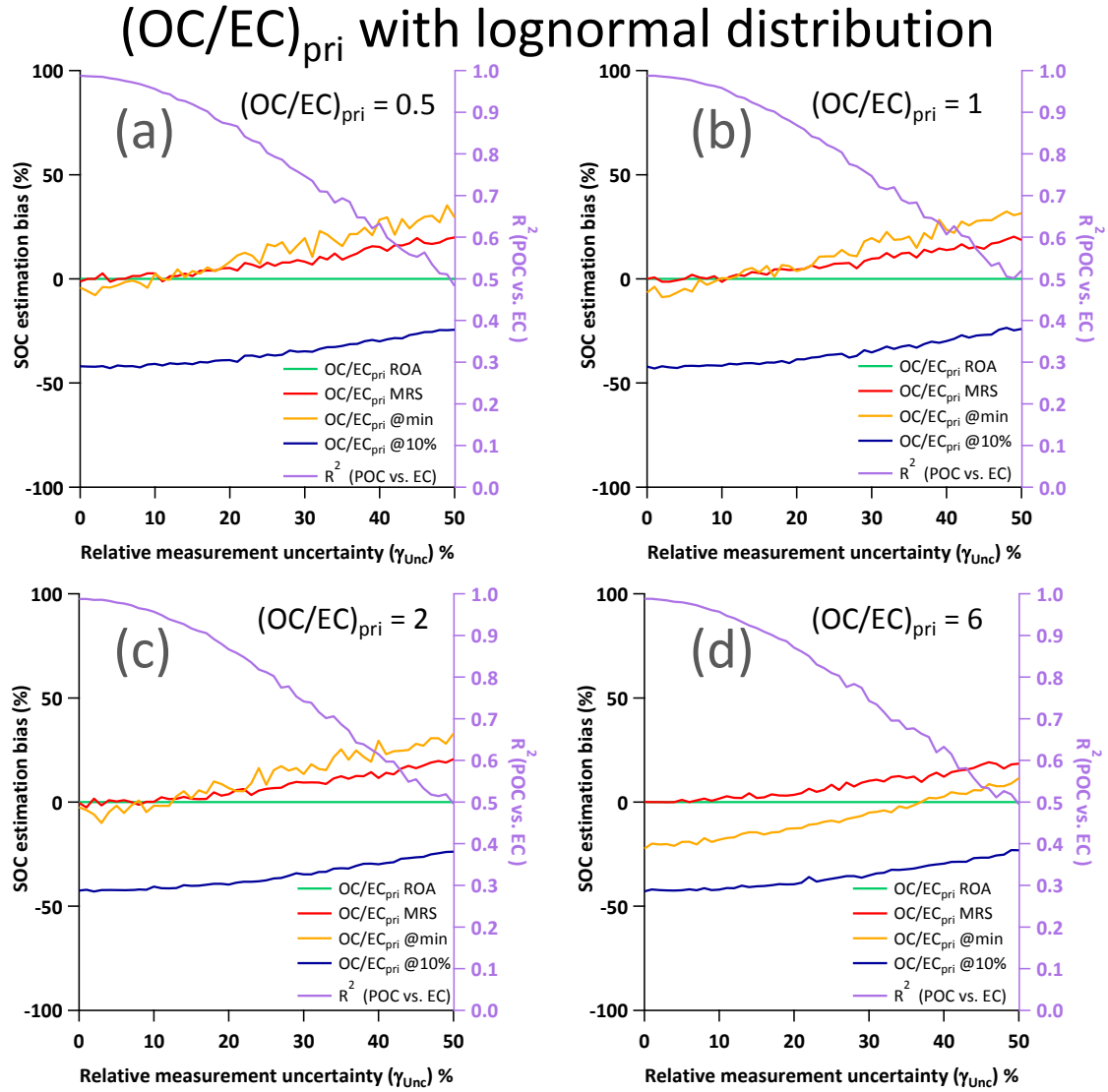
Figure S13. Bias of SOC determination as a function of measurement uncertainty and (OC/EC)$_{pri}$ by different representation of (OC/EC)$_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), OC/EC$_{10\%}$, OC/EC$_{min}$. Fixed input parameters: N=8000; EC = 2±1 µgm^{-3}, and f$_{SOC}$=60%. (OC/EC)$_{pri}$ is varied, with (a) (OC/EC)$_{pri}$ = 0.5±0.025, (b) (OC/EC)$_{pri}$ = 1±0.05, (c) (OC/EC)$_{pri}$ = 2 ±0.1, and (d) (OC/EC)$_{pri}$ = 6±0.3

$(OC/EC)_{pri}$ with lognormal distribution
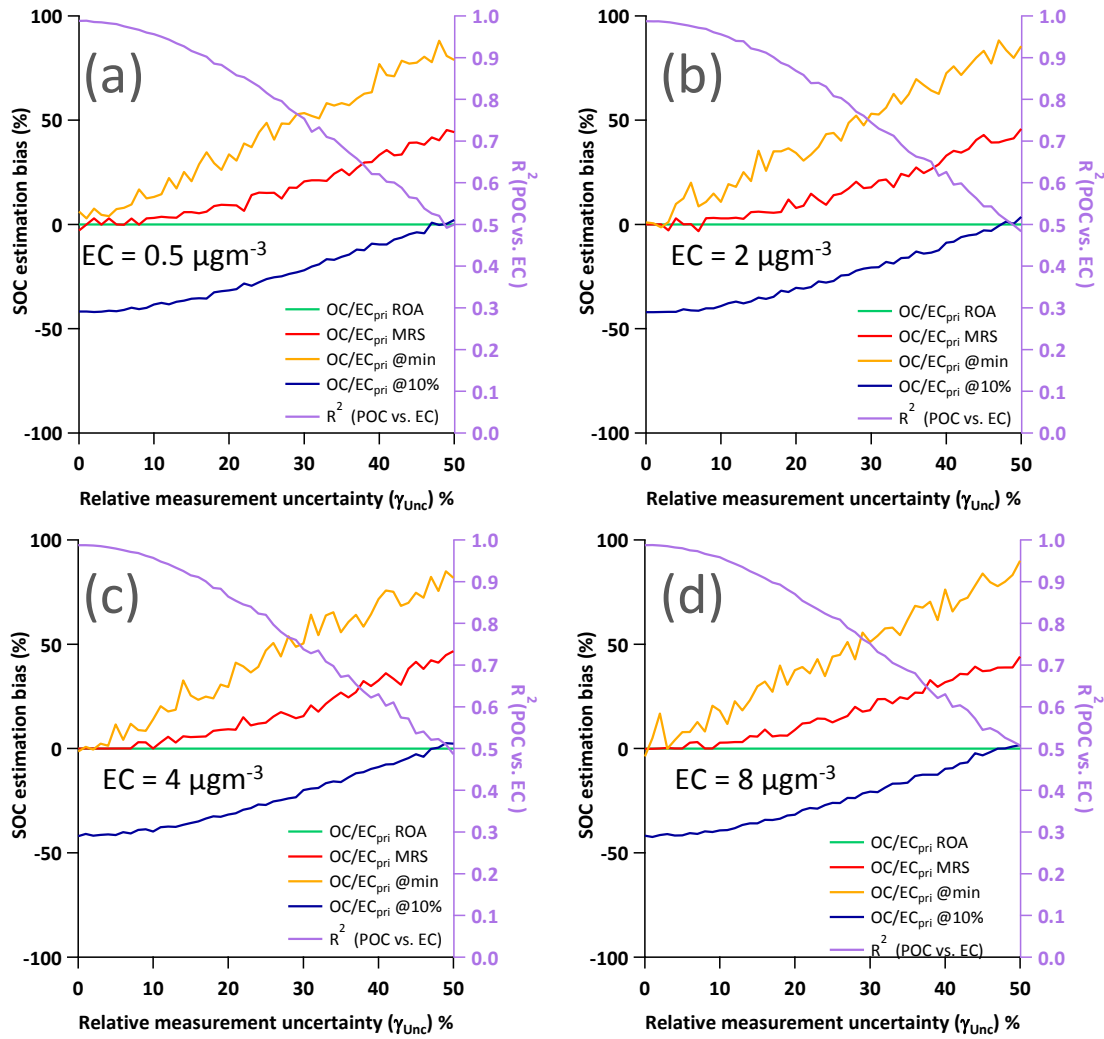
Figure S14. Bias of SOC determination as a function of measurement uncertainty and EC by different representation of $(OC/EC)_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), $OC/EC_{10\%}$ and OC/EC_{min}. Fixed input parameters: N=8000; $(OC/EC)_{pri}$ = 0.5±0.025; and f_{SOC} =40%. EC and SOC input configurations are varied, with (a) EC = 0.5±0.25 μgm^{-3}, SOC = 0.167±0.083 μgm^{-3}, (b) EC = 2±1 μgm^{-3}, SOC = 0.67±0.33 μgm^{-3}, (c) EC = 4±2 μgm^{-3}, SOC = 1.32±0.66 μgm^{-3}, and (d) EC = 8±4 μgm^{-3}, SOC = 2.64±1.32 μgm^{-3}.

3. An Igor program for MRS calculation

An Igor Pro (WaveMetrics, Inc. Lake Oswego, OR, USA) based program with graphical user interface is developed to make the calculation feasible and user friendly (Figure S15). Data input can be easily accomplished through copy and paste from Excel directly. MRS calculation on subset of data is also available (for example, monthly, hourly and seasonal subset data). The program can be downloaded from https://sites.google.com/site/wuchengust.
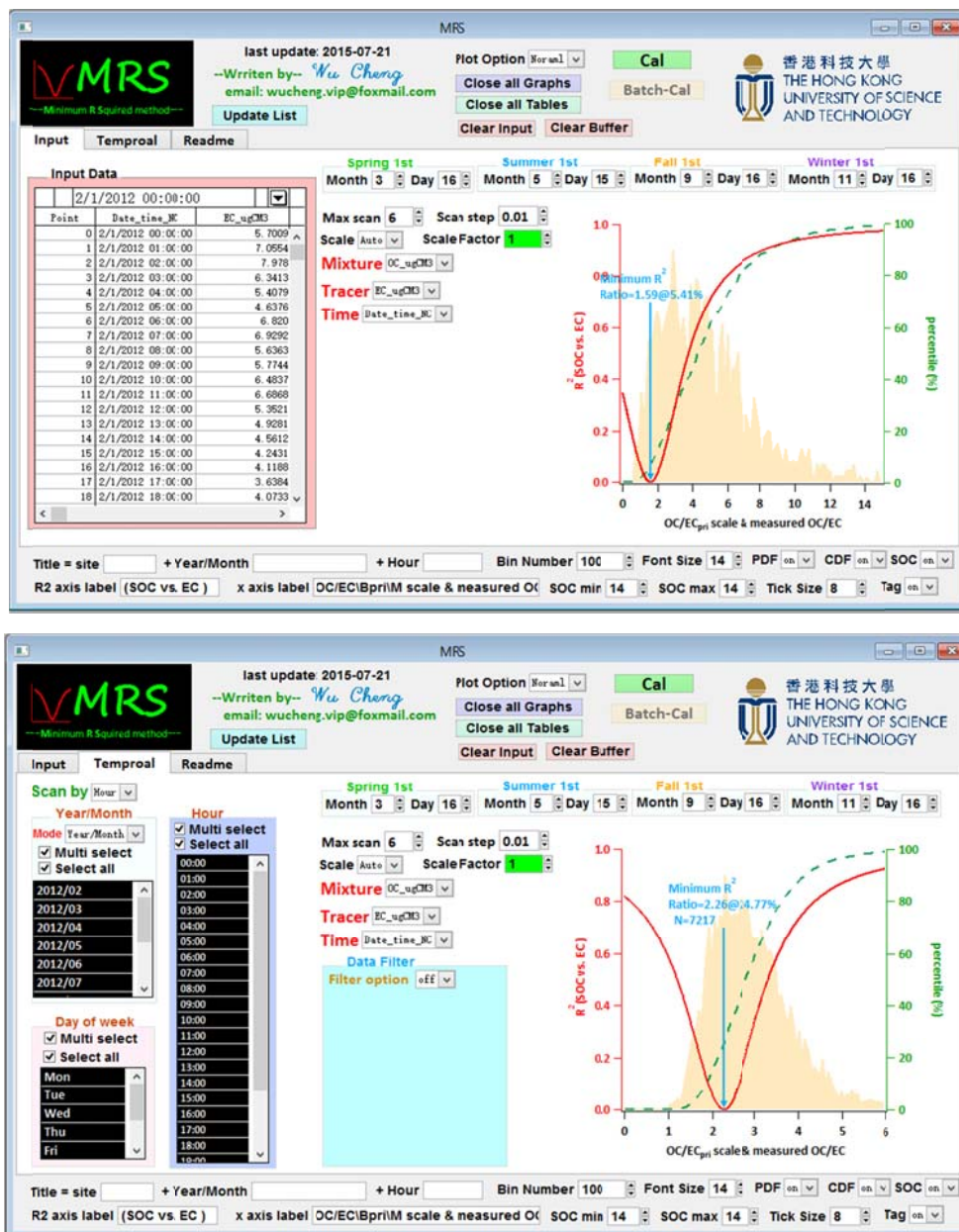
Figure S15. MRS Igor program. The calculation can be done on different time scale including hourly, monthly and seasonally. MRS can also be calculated on grouped data based on a text data column provided by the user.