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)

S1. 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.

S2. 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 S9 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 S10. 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}. Results of sensitivity tests are discussed below are summarized in Table S2 considering most possible ambient conditions.

As shown in Figure S11, $OC/EC_{10\%}$ 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 S13. 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 ($\gamma_{Unc} < 20\%$), 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 S14 and S15. 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 S14d).

As shown in Figure S17, the sample size dependency of all three $(OC/EC)_{pri}$ representatives is not sensitive to f_{SOC} and these results are similar to the results with single $(OC/EC)_{pri}$ (Figure S16). For the sample size dependency, the log-normally distributed $(OC/EC)_{pri}$ (Figure S18) behave similar to the single value $(OC/EC)_{pri}$ (Figure 8). When EC concentration is higher than 2 μgC m^{-3}, case B (absolute measurement uncertainty 0.2 μgC m^{-3}) is equivalent to case A (relative measurement uncertainty 10%) with a Υ_{unc} lower than 10%. As a result, the SOC bias of MRS in case A (Figure S18a and Table S3) is slightly higher than case B (Figure S18b and Table S4).

S3. Igor programs for MRS calculation and simulated OC EC data generation using MT

An Igor Pro (WaveMetrics, Inc. Lake Oswego, OR, USA) based program with graphical user interface (GUI) is developed to make the calculation feasible and user friendly (Figure S19). 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).

Another program using MT can generate simulated OC and EC concentration through user defined parameters via GUI as shown in Figure S20.

Both programs can be downloaded from https://sites.google.com/site/wuchengust.

References

Chu, S. H.: Stable estimate of primary OC/EC ratios in the EC tracer method, Atmos Environ, 39, 1383-1392, 2005.

Saylor, R. D., Edgerton, E. S., and Hartsell, B. E.: Linear regression techniques for use in the EC tracer method of secondary organic aerosol estimation, Atmos Environ, 40, 7546-7556, DOI 10.1016/j.atmosenv.2006.07.018, 2006.

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.

		SOC bias due to tested parameter		
	Tested parameter	MRS	OC/EC_{min}	$OC/EC_{10\%}$
Scenario 1 Single source	RSD_{EC}	±4%	±20%	-45% ~ -40%
	RSD_{SOC}	±4%	-20% ~ 0%	-50% ~ -20%
	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%;

Table S3. Data used in Figure S18a. Bias of SOC determination as a function of sample size by different representation of $(OC/EC)_{pri}$, including: minimum R square method (MRS), $OC/EC_{10\%}$ and OC/EC_{min}. For each sample size, 500 repeat runs were conducted. Fixed input parameters: n = 20~8000; EC = 8±4 μgC m^{-3}, f_{SOC} =40% and $(OC/EC)_{pri}$ = 0.5±0.025. Measurement uncertainty：

Υ_{unc}= 10%.

Sample size	MRS		$OCEC_{10\%}$		$OCEC_{min}$	
	Mean	S.D.*	Mean	S.D.*	Mean	S.D.*
20	2.20%	28.44%	-39.73%	9.16%	-22.09%	11.71%
30	2.01%	21.27%	-40.01%	8.05%	-18.00%	9.97%
40	2.29%	18.65%	-39.57%	7.10%	-15.67%	9.81%
50	3.09%	15.60%	-39.27%	6.29%	-13.29%	9.36%
60	2.34%	15.36%	-39.43%	5.65%	-12.51%	8.55%
70	3.08%	13.75%	-39.17%	5.28%	-10.02%	8.55%
80	3.27%	13.41%	-39.14%	4.85%	-9.70%	7.98%
90	1.69%	11.65%	-39.01%	4.86%	-8.58%	7.68%
100	2.11%	10.84%	-39.05%	4.09%	-7.85%	7.75%
200	2.27%	8.26%	-38.95%	3.26%	-3.56%	6.63%
300	2.21%	6.65%	-38.89%	2.68%	-1.74%	6.03%
400	2.48%	5.87%	-38.72%	2.24%	-0.30%	5.28%
500	2.69%	5.28%	-38.84%	2.00%	0.93%	4.90%
600	2.40%	4.58%	-38.77%	1.88%	1.63%	4.59%
700	2.24%	4.39%	-38.78%	1.74%	2.36%	4.41%
800	2.30%	3.95%	-38.82%	1.53%	2.24%	4.39%
900	2.44%	4.10%	-38.68%	1.63%	3.03%	4.14%
1000	2.49%	3.69%	-38.76%	1.48%	3.07%	4.27%
2000	2.34%	2.61%	-38.81%	1.01%	3.20%	3.00%
3000	2.51%	2.16%	-38.69%	0.80%	3.10%	2.57%
4000	2.34%	1.92%	-38.77%	0.70%	3.39%	2.15%
5000	2.46%	1.78%	-38.75%	0.64%	3.42%	1.93%
6000	2.53%	1.72%	-38.74%	0.61%	3.29%	1.77%
7000	2.46%	1.57%	-38.73%	0.54%	3.21%	1.64%
8000	2.35%	1.48%	-38.73%	0.53%	3.31%	1.53%

* S.D.: standard deviation

Table S4. Data used in Figure S18b. Bias of SOC determination as a function of sample size by different representation of $(OC/EC)_{pri}$, including: minimum R square method (MRS), $OC/EC_{10\%}$ and OC/EC_{min}. For each sample size, 500 repeat runs were conducted. Fixed input parameters: n = 20~8000; EC = 8±4 μgC m^{-3}, f_{SOC} = 40% and $(OC/EC)_{pri}$ = 0.5±0.025. Measurement uncertainty: $\varepsilon_{EC}, \varepsilon_{OC}$= **0.2 μgC m^{-3}**.

Sample size	MRS		$OCEC_{10\%}$		$OCEC_{min}$	
	Mean	S.D.*	Mean	S.D.*	Mean	S.D.*
20	2.65%	26.58%	-42.44%	8.18%	-28.38%	9.35%
30	1.13%	19.92%	-42.37%	7.31%	-25.57%	8.39%
40	-0.41%	17.42%	-42.29%	6.04%	-23.30%	7.43%
50	0.74%	14.82%	-42.66%	5.46%	-21.16%	6.87%
60	-0.31%	14.35%	-42.24%	5.02%	-20.31%	6.37%
70	-0.60%	12.28%	-42.22%	4.88%	-19.36%	5.99%
80	0.33%	11.89%	-42.29%	4.39%	-19.35%	5.68%
90	0.12%	11.39%	-42.38%	4.36%	-18.56%	5.95%
100	-0.26%	11.05%	-42.61%	4.12%	-17.80%	5.77%
200	0.54%	7.80%	-42.26%	2.84%	-14.60%	4.68%
300	0.16%	5.95%	-41.96%	2.42%	-13.28%	4.13%
400	-0.08%	5.09%	-42.04%	2.11%	-12.58%	3.68%
500	0.53%	4.79%	-41.85%	1.91%	-11.87%	3.44%
600	0.17%	4.09%	-42.17%	1.59%	-11.33%	3.36%
700	0.05%	3.76%	-42.08%	1.47%	-10.82%	3.17%
800	-0.07%	3.55%	-41.99%	1.36%	-10.67%	2.84%
900	-0.05%	3.64%	-42.03%	1.34%	-10.36%	2.97%
1000	0.03%	3.18%	-42.01%	1.28%	-10.07%	2.60%
2000	0.11%	2.49%	-41.93%	0.88%	-10.00%	1.83%
3000	0.13%	1.98%	-41.94%	0.73%	-10.07%	1.60%
4000	0.12%	1.86%	-41.87%	0.63%	-10.04%	1.53%
5000	0.09%	1.71%	-41.93%	0.59%	-10.00%	1.28%
6000	0.05%	1.39%	-41.98%	0.52%	-10.10%	1.20%
7000	0.01%	1.48%	-41.95%	0.48%	-9.93%	1.08%
8000	0.13%	1.27%	-41.94%	0.46%	-9.98%	1.07%

* S.D.: standard deviation

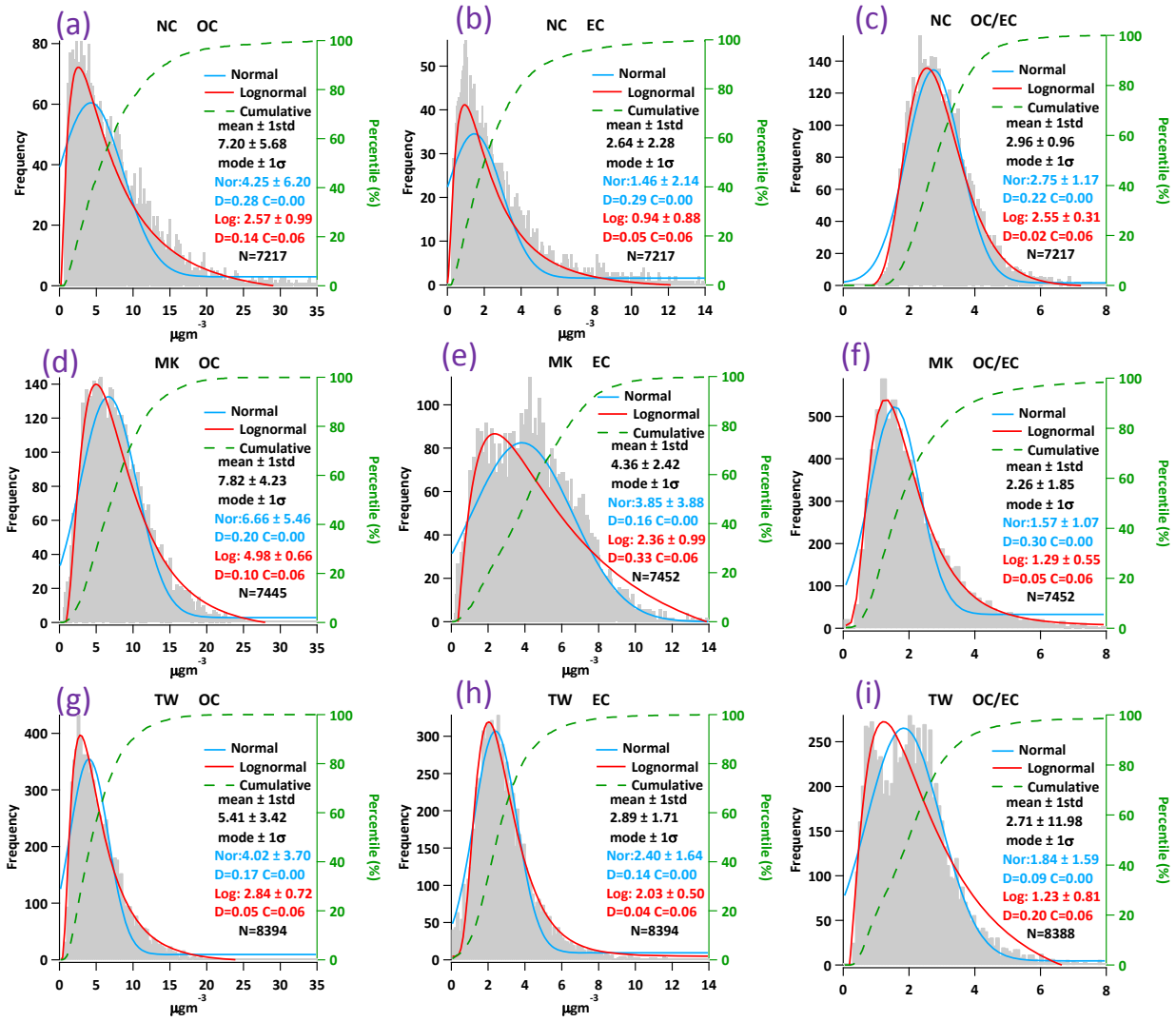


Figure S1. Frequency distributions of OC, EC and OC/EC ratio in ambient samples recorded at three locations in the Pearl River Delta region, China. (a)-(c) Nancun, a suburban site in Guangzhou, (d)-(f) Mong Kok, an urban roadside site in Hong Kong, (g)-(i) Tsuen Wan, a general urban site in Hong Kong. Grey areas represent frequency distribution, green dashed lines represent cumulative frequency distribution, blue lines are normal fitting, and red lines are log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.

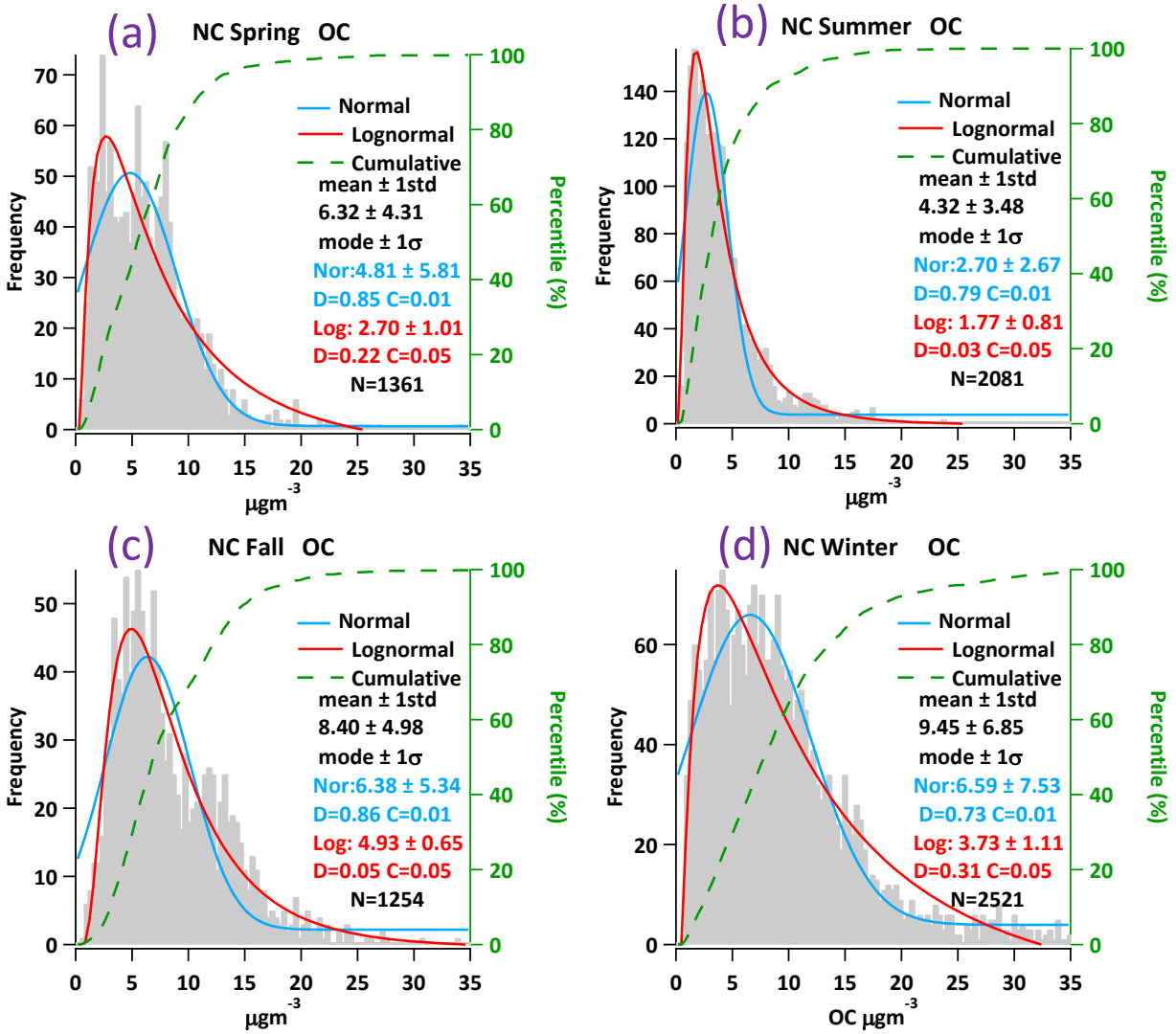


Figure S2. Seasonal frequency distributions of **OC** observed at the suburban site Nancun in the Pearl River Delta region. (a)-(d) represent spring, summer, fall, and winter, respectively. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, blue lines represent normal fitting, red lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.

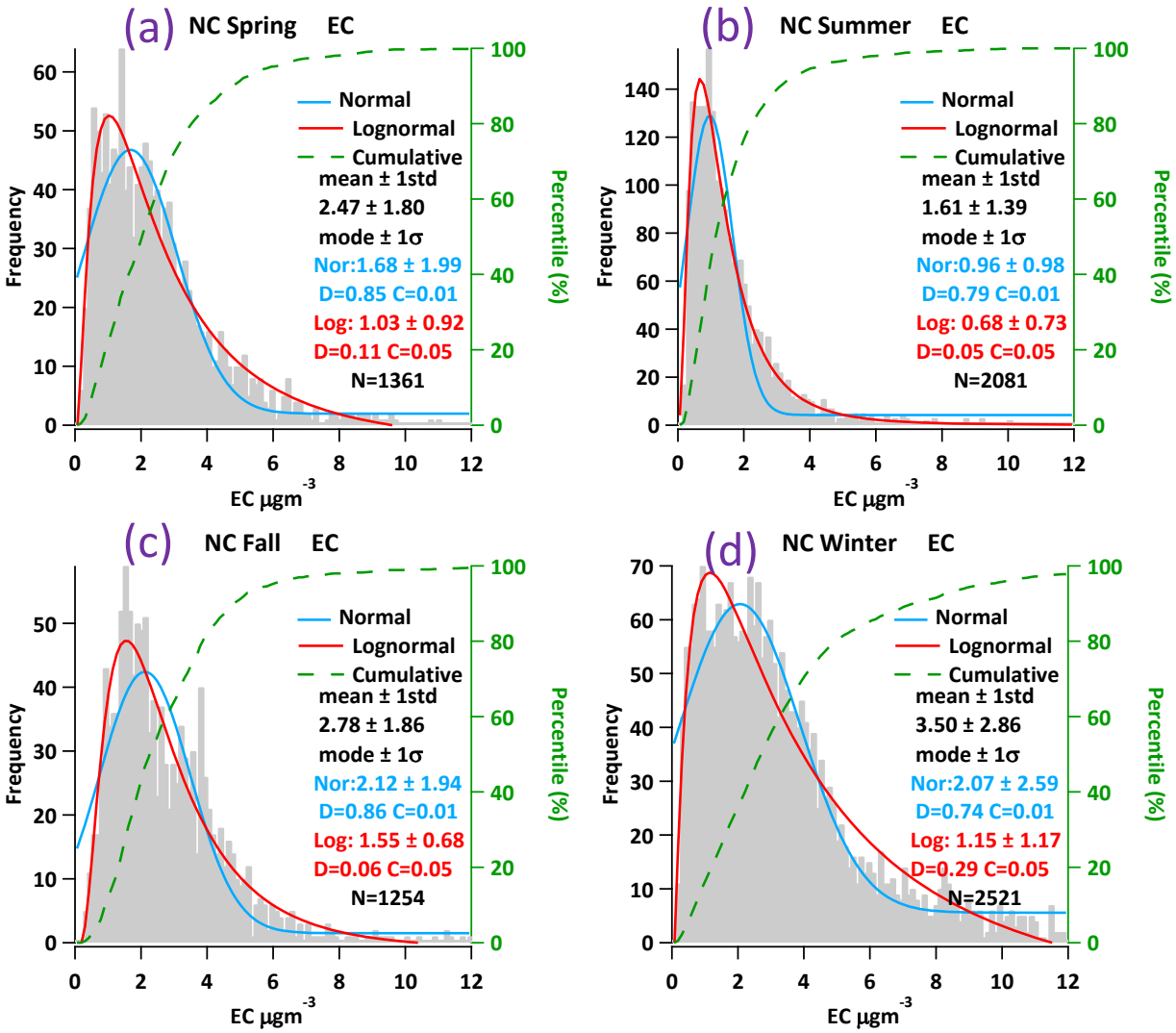


Figure S3. Seasonal frequency distributions of **EC** observed at the suburban site Nancun in the Pearl River Delta region. (a)-(d) represent spring, summer, fall, and winter, respectively. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, blue lines represent normal fitting, red lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.

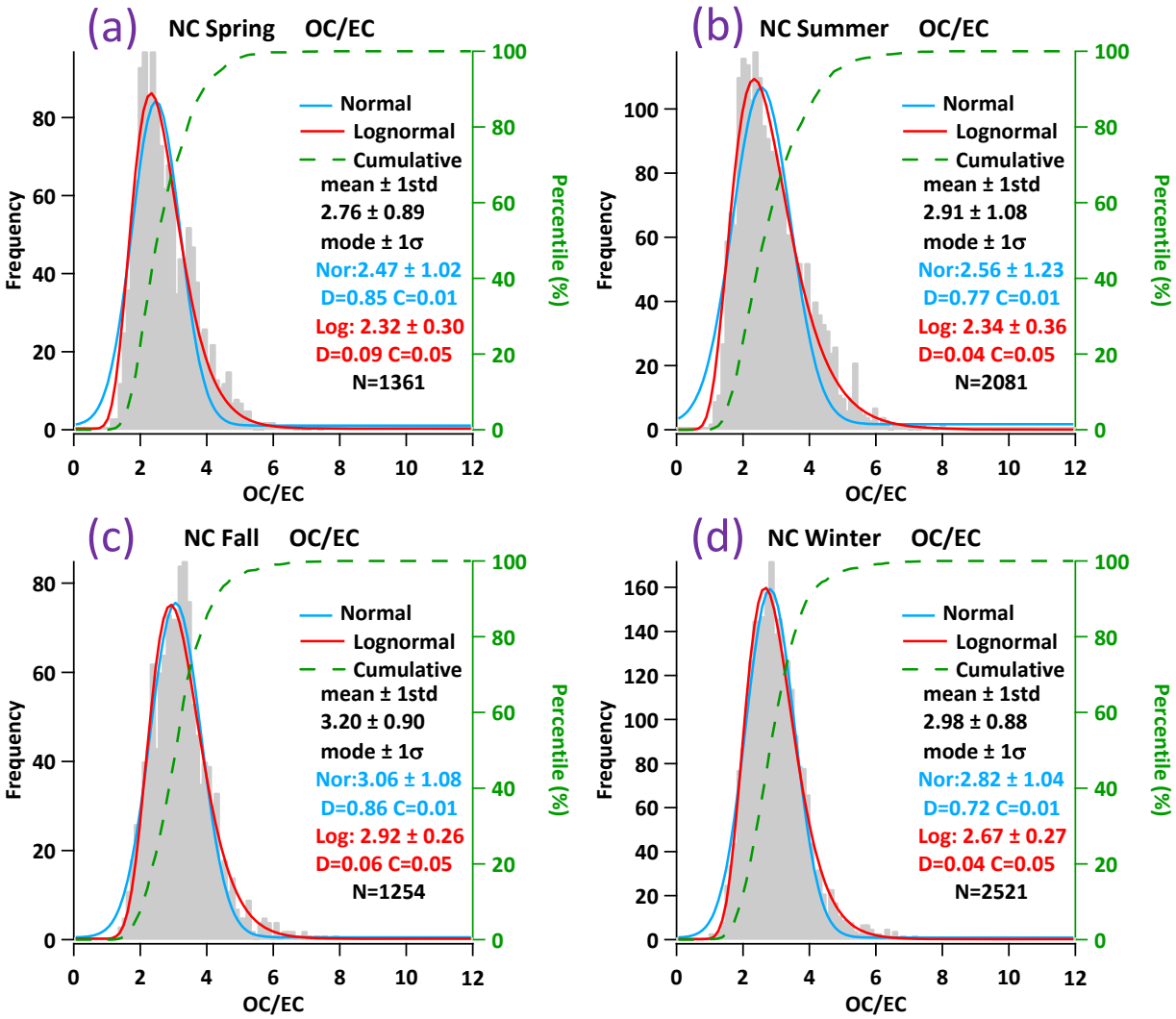


Figure S4. Seasonal frequency distributions of **OC/EC** ratio observed at the suburban site Nancun in the Pearl River Delta region. (a)-(d) represent spring, summer, fall, and winter, respectively. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, blue lines represent normal fitting, red lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.

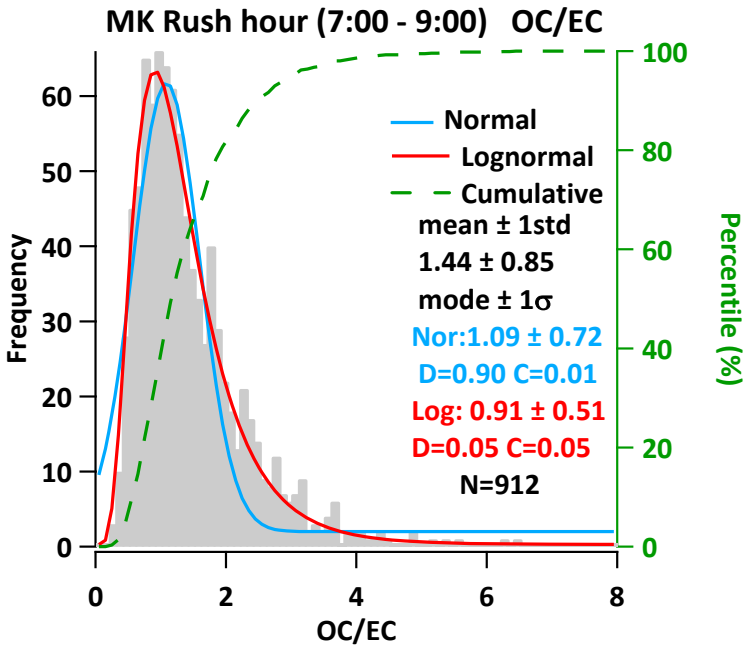


Figure S5. Frequency distribution of OC/EC ratio observed at Mong Kok (MK), a roadside site in Hong Kong. OC/EC ratio of road site MK can represent the $(OC/EC)_{pri}$ in Hong Kong. Grey areas represent frequency distributions, green dashed lines represent cumulative frequency distributions, blue lines represent normal fitting, red lines represent log-normal fitting. D represent the Kolmogorov–Smirnov statistic, C represent critical value. If D<C, the samples follow corresponding distribution.

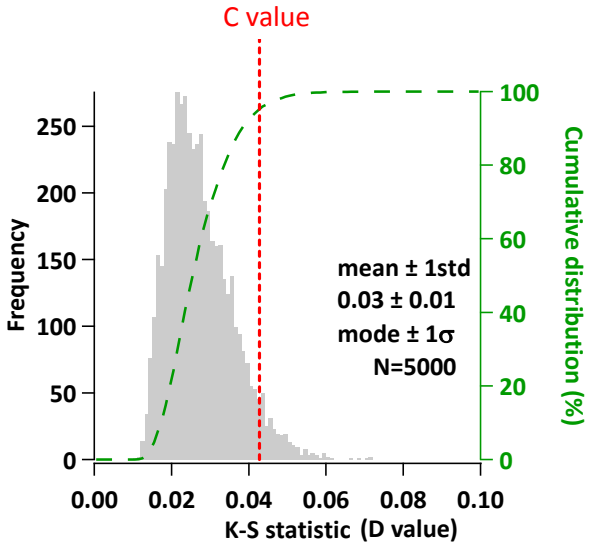


Figure S6. Performance of the MT pseudorandom number generator evaluated by K-S test. The histogram in grey represents D statistic value in K-S test and the red dashed-line represents C. The dash line in green represents cumulative distribution of D. Data with D<C, i.e., data that strictly follow the log-normal distribution, account for 94.4% in 5000 runs.

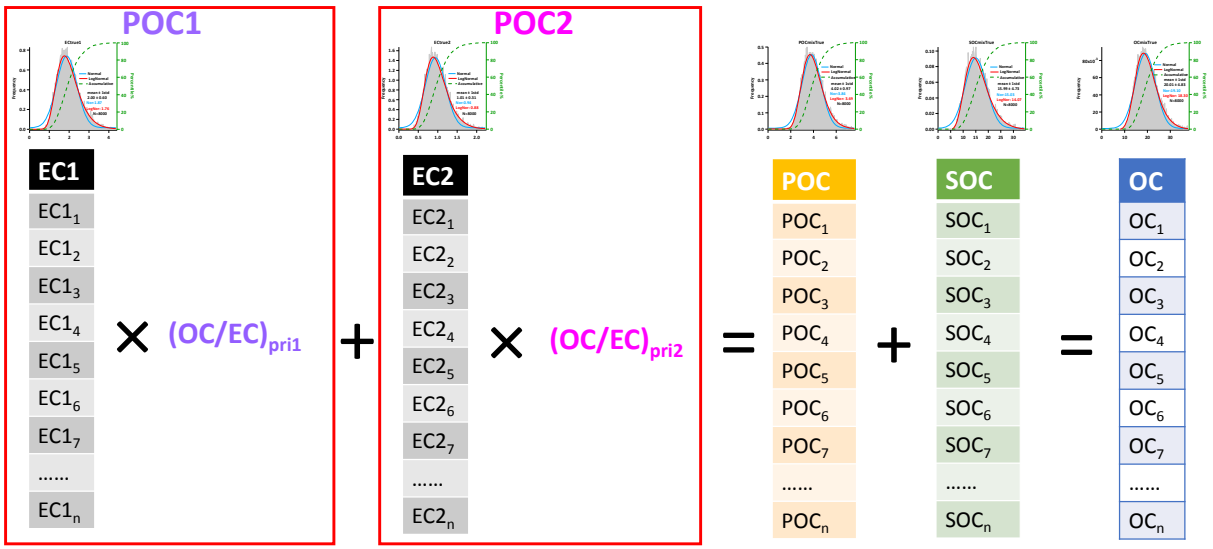


Figure S7. Schematic diagram of pseudorandom number generation for two primary emission sources. The data series (EC_1, EC_2 and SOC), which are generated by Mersenne twister (MT) pseudorandom number generator, statistically follow log-normal distribution, but the sequence of each data point is randomly assigned.

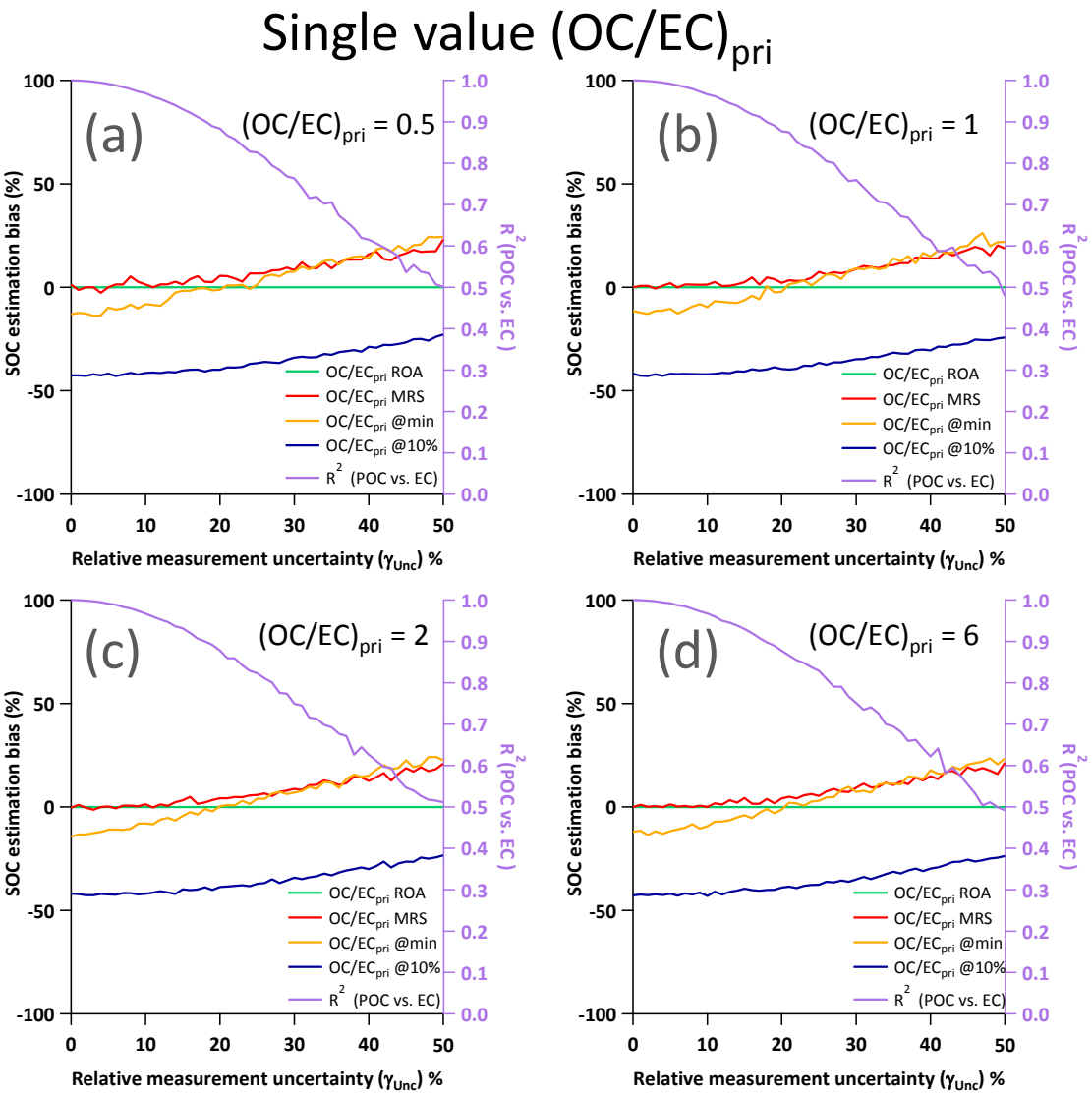


Figure S8. Variation of bias of SOC determination with measurement uncertainty under four different $(OC/EC)_{pri}$ (a) $(OC/EC)_{pri}$ = 0.5, (b) $(OC/EC)_{pri}$ = 1, (c) $(OC/EC)_{pri}$ = 2 , and (d) $(OC/EC)_{pri}$ = 6, demonstrating dependence on γ_{unc} while no effects by $(OC/EC)_{pri}$. The different 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 µgC m^{-3}, and f_{SOC} = 60%.

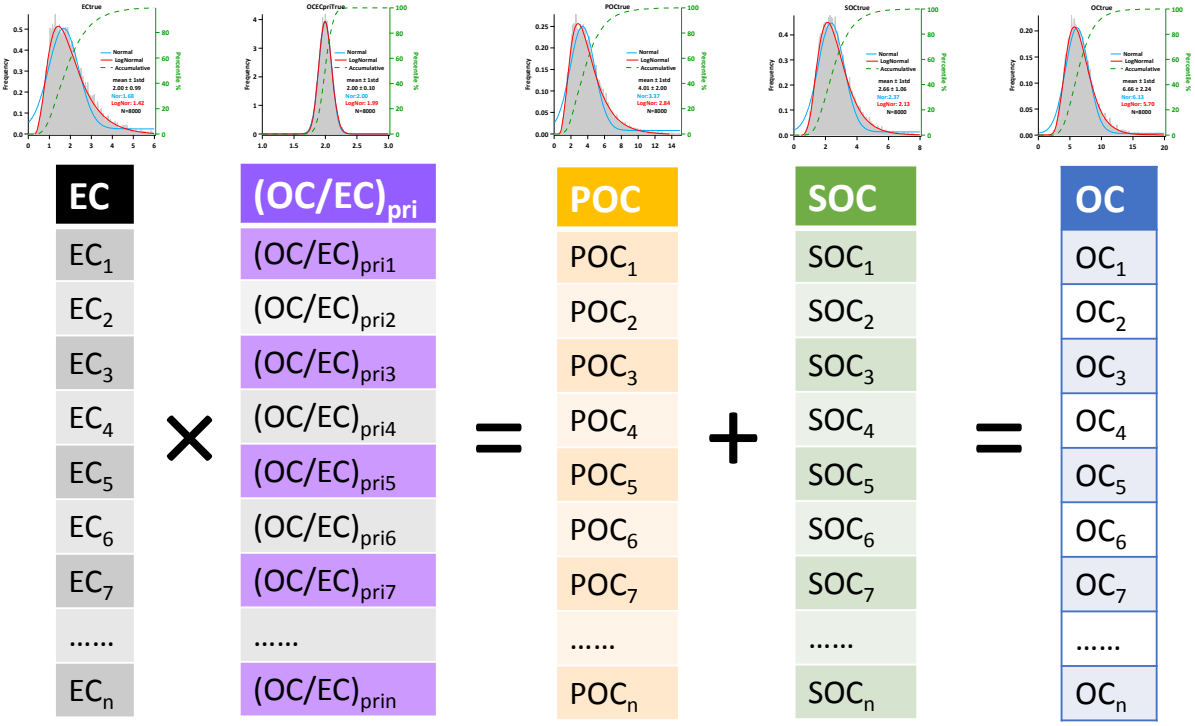


Figure S9. Schematic diagram of pseudorandom number generation. The data series (EC, $(OC/EC)_{pri}$ and SOC), which are generated by Mersenne twister (MT) pseudorandom number generator, statistically follow log-normal distribution, but the sequence of each data point is randomly assigned.

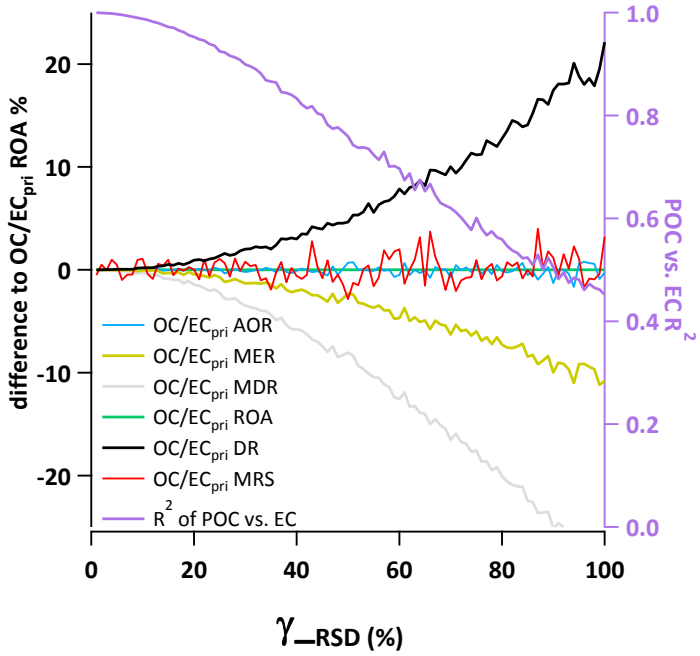


Figure S10. Deviation of various representation of $(OC/EC)_{pri}$ from the ratio of averages (ROA) as a function of $(OC/EC)_{pri}$ RSD to EC RSD ratio ($\gamma_{\_RSD}$). Representation of $(OC/EC)_{pri}$ including: average of ratios(AOR), median of ratios(MER), mode of ratios(MDR), deming regression (DR), minimum R square method (MRS). Input configurations: n = 8000, EC = 4±2 μgC m^{-3}, $(OC/EC)_{pri}$ = 2, SOC/OC ratio (f_{SOC}) = 40% and SOC = 5.3±2.6 μgC m^{-3}.

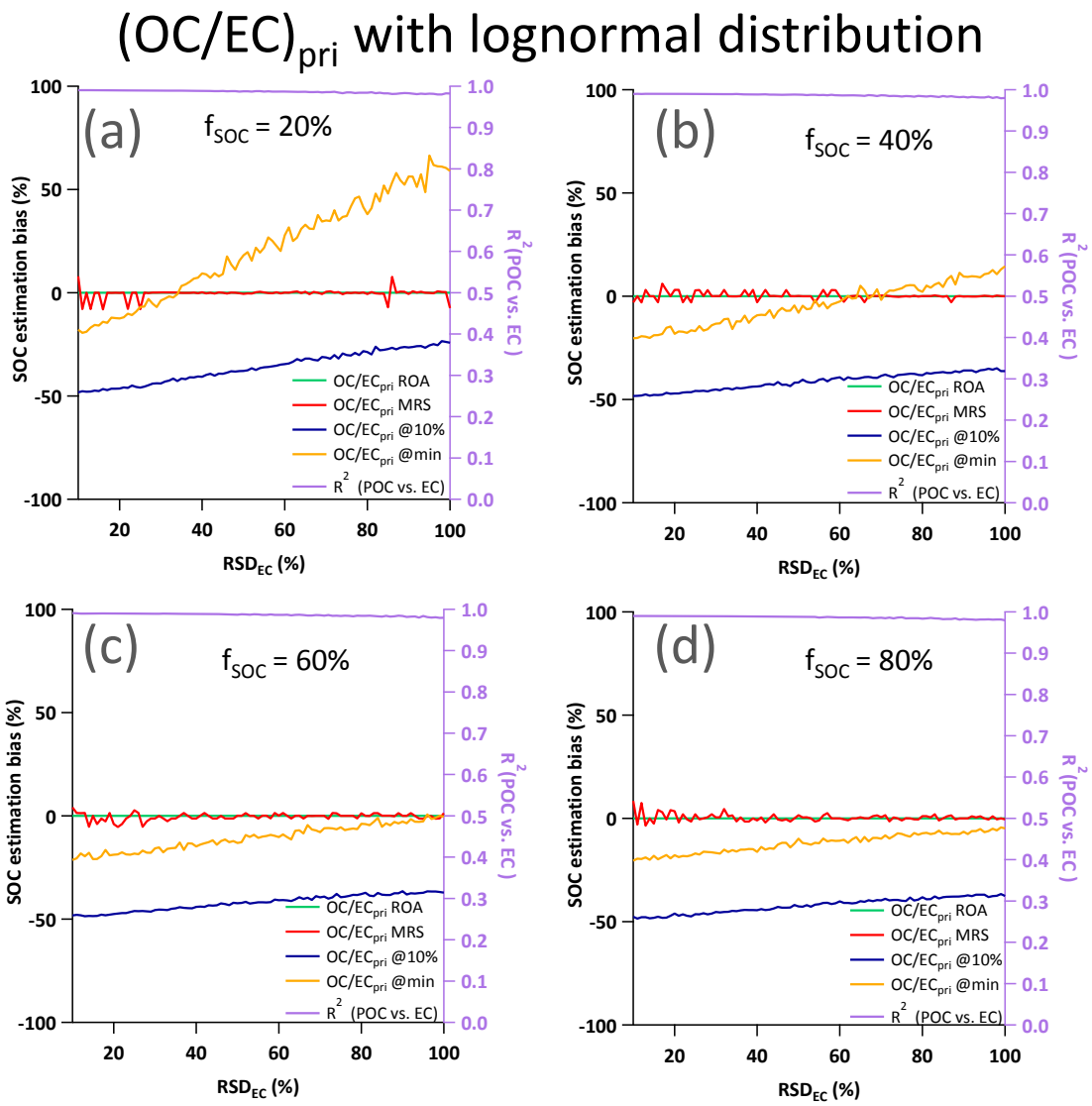


Figure S11. Bias of SOC determination as a function of RSD_{EC} and f_{SOC} by different representation of $(OC/EC)_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), $OC/EC_{10\%}$, OC/EC_{min}. Fixed input parameters: n = 8000, EC = 2±1 μgC m^{-3}, and $(OC/EC)_{pri}$ = 0.5±0.025. f_{SOC} is varied, with (a) f_{SOC} = 20% SOC = 0.25±0.13 μgC m^{-3} (b) f_{SOC} =40% SOC = 0.67±0.33 μgC m^{-3} (c) f_{SOC} =60% SOC = 1.5±0.75 μgC m^{-3} (d) f_{SOC} =80% SOC = 4±2 μgC m^{-3}

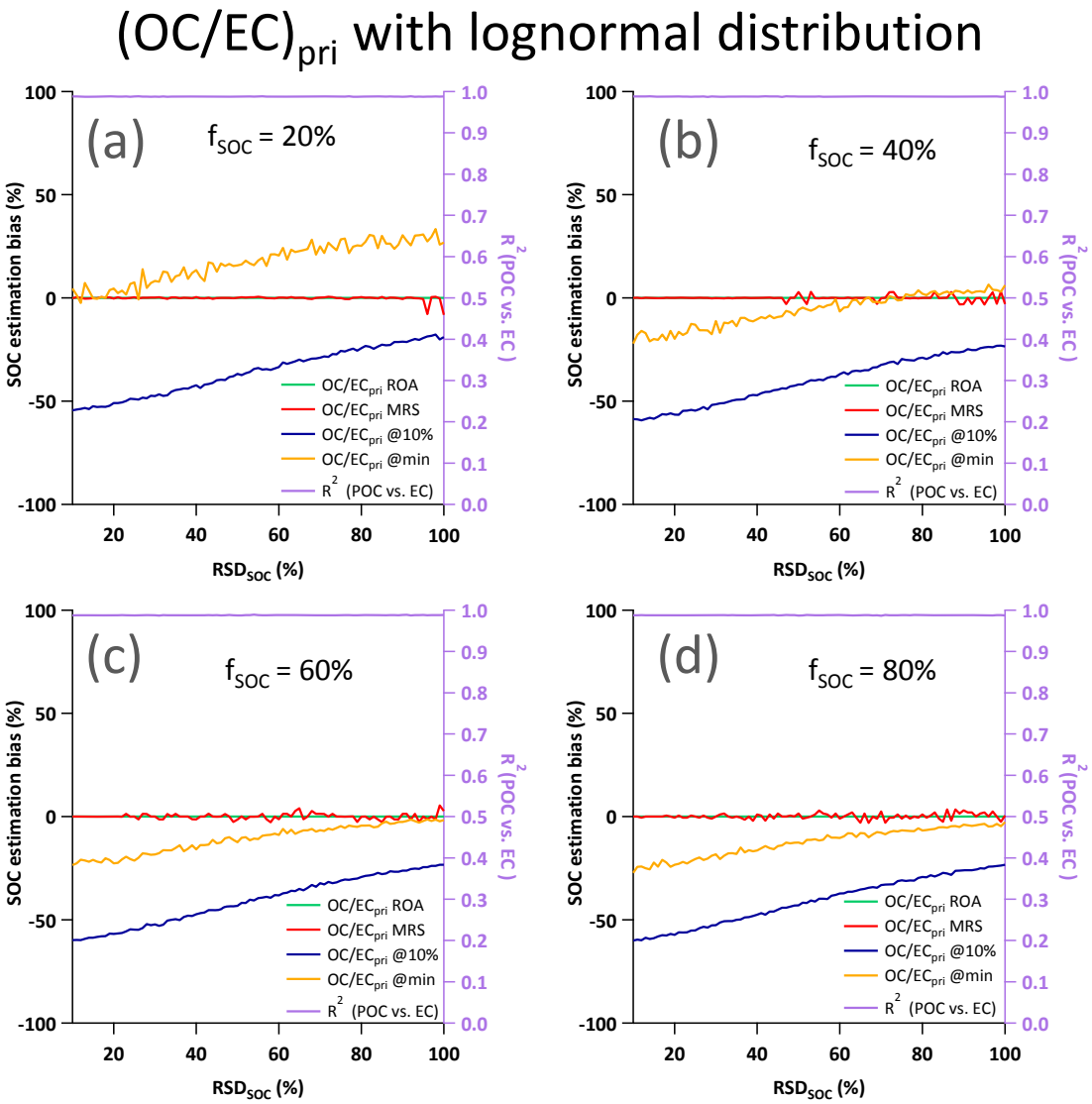


Figure S12. Bias of SOC determination as a function of RSD_{SOC} and f_{SOC} by different representation of $(OC/EC)_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), $OC/EC_{10\%}$, OC/EC_{min}. Fixed input parameters: n = 8000, EC = 2±1 μgC m^{-3}, and $(OC/EC)_{pri}$ = 0.5±0.025.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

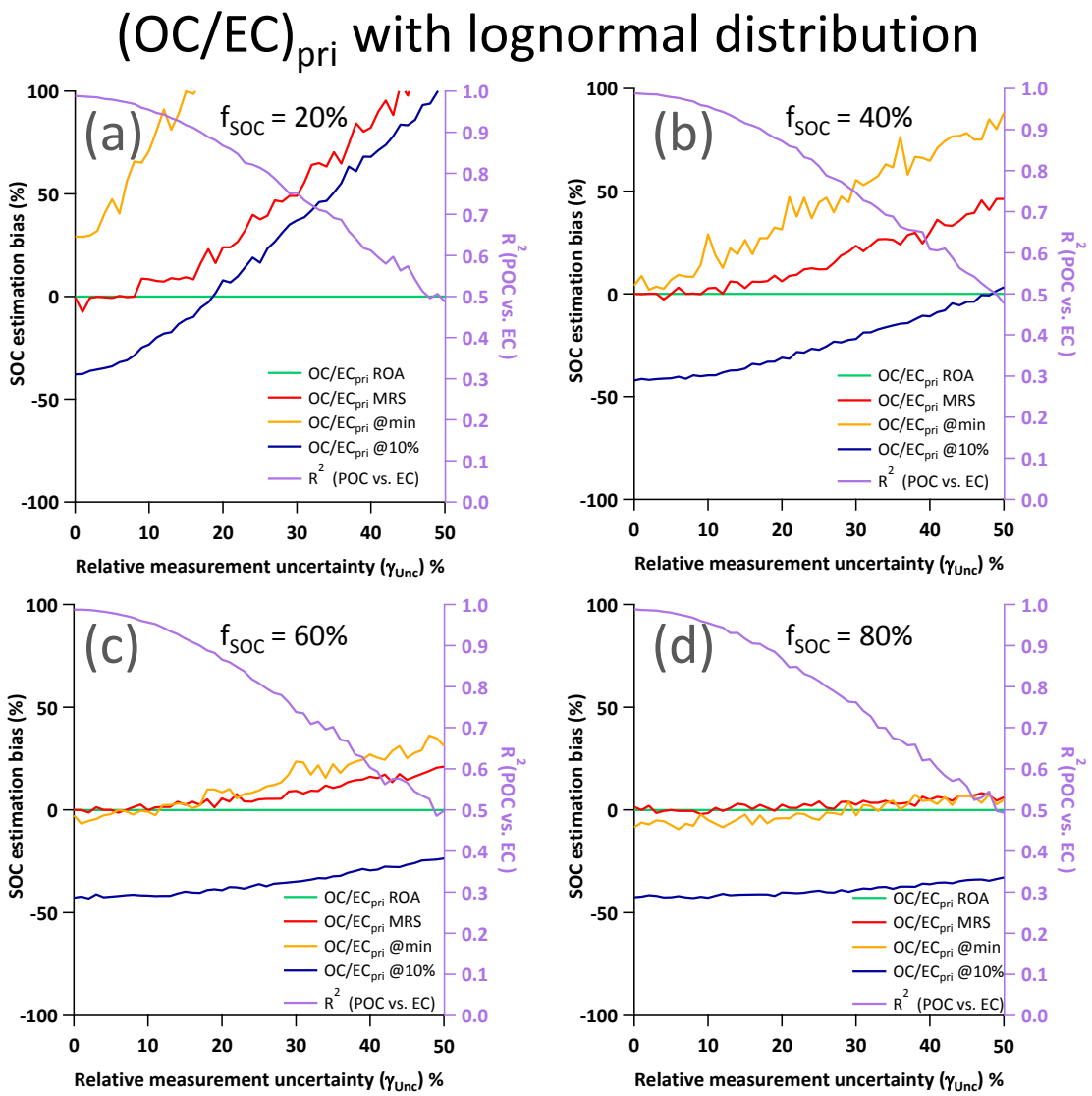


Figure S13. Bias of SOC determination as a function of relative measurement uncertainty (γ_{Unc}) and $(OC/EC)_{pri}$ by different representation of $(OC/EC)_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), $OC/EC_{10\%}$, OC/EC_{min}. Fixed input parameters: n = 8000, EC = 2±1 μgC m^{-3} and $(OC/EC)_{pri}$ = 0.5±0.025. f_{SOC} is varied, with (a) f_{SOC} = 20% SOC = 0.25±0.13 μgC m^{-3} (b) f_{SOC} = 40% SOC = 0.67±0.33 μgC m^{-3} (c) f_{SOC} = 60% SOC = 1.5±0. 75 μgC m^{-3} (d) f_{SOC} = 80% SOC = 4±2 μgC m^{-3}

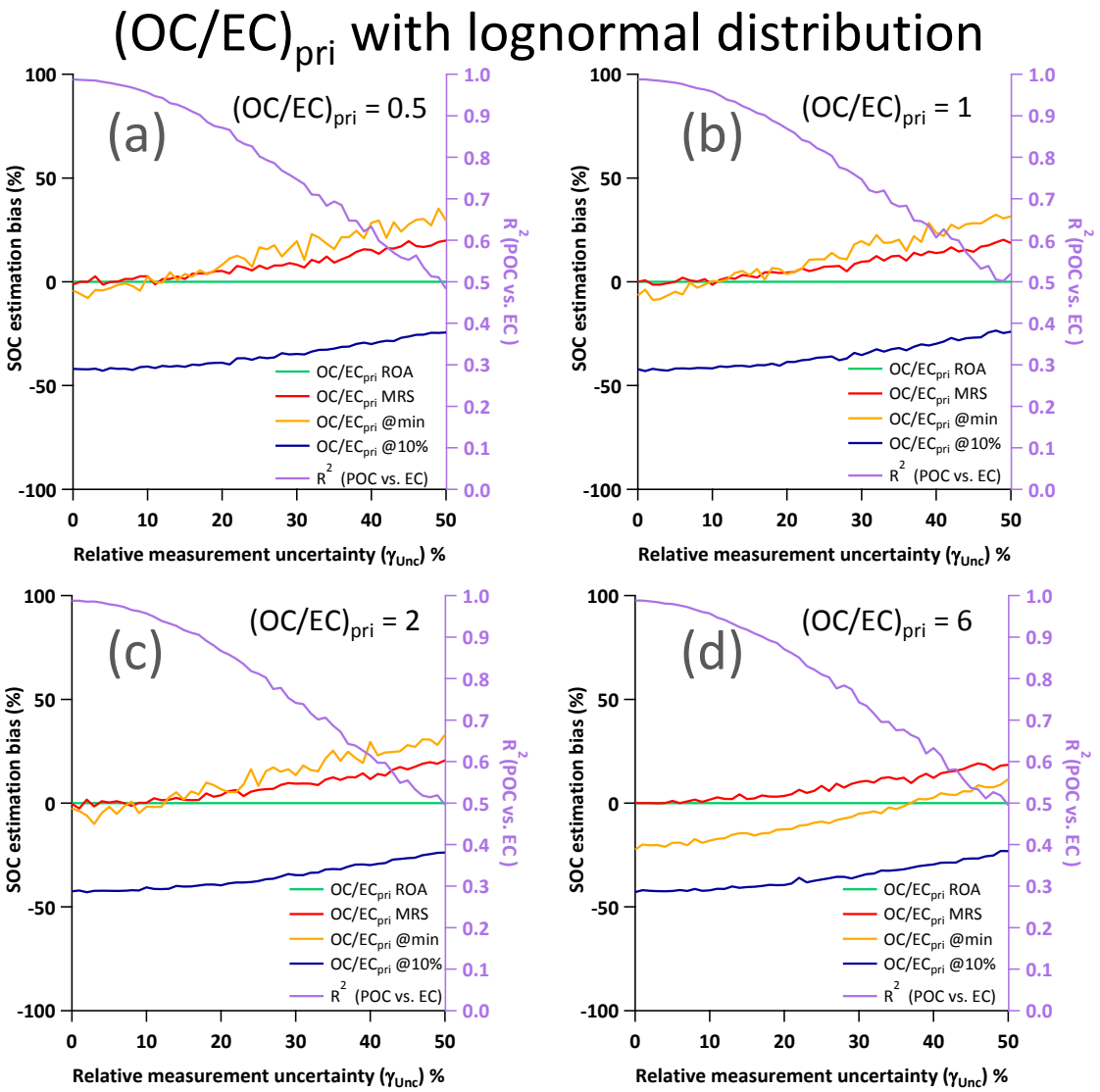


Figure S14. Bias of SOC determination as a function of measurement uncertainty and $(OC/EC)_{pri}$ by different representation of $(OC/EC)_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), $OC/EC_{10\%}$, OC/EC_{min}. Fixed input parameters: n = 8000, EC = 2±1 μgC m^{-3} and 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

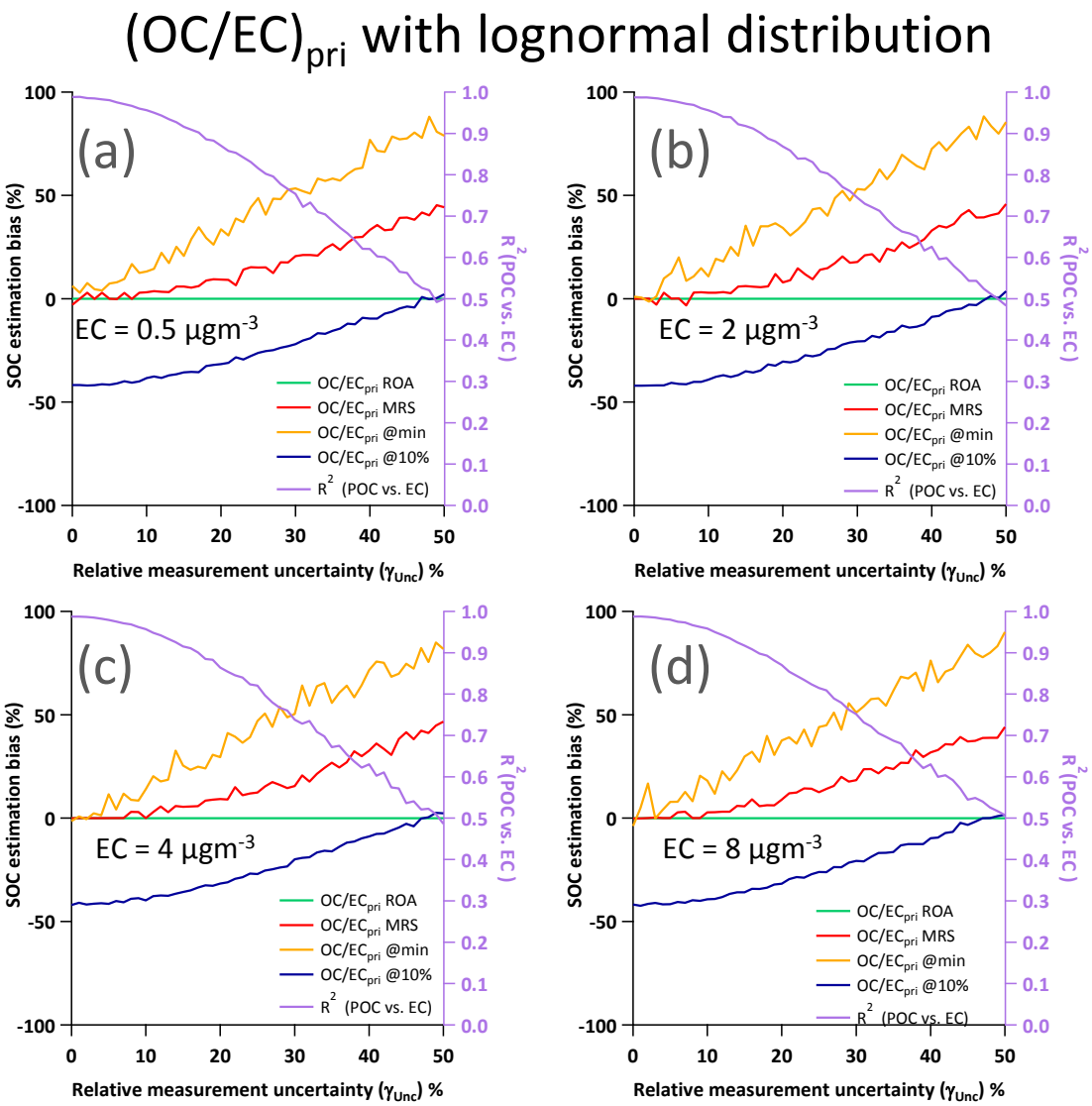


Figure S15. Bias of SOC determination as a function of measurement uncertainty and EC by different representation of $(OC/EC)_{pri}$, including: ratio of averages (ROA), minimum R square method (MRS), $OC/EC_{10\%}$ and OC/EC_{min}. Fixed input parameters: n = 8000, $(OC/EC)_{pri}$ = 0.5±0.025 and f_{SOC} =40%. EC and SOC input configurations are varied, with (a) EC = 0.5±0.25 µgC m^{-3}, SOC = 0.167±0.083 µgC m^{-3}, (b) EC = 2±1 µgC m^{-3}, SOC = 0.67±0.33 µgC m^{-3}, (c) EC = 4±2 µgC m^{-3}, SOC = 1.32±0.66 µgC m^{-3}, and (d) EC = 8±4 µgC m^{-3}, SOC = 2.64±1.32 µgC m^{-3}.

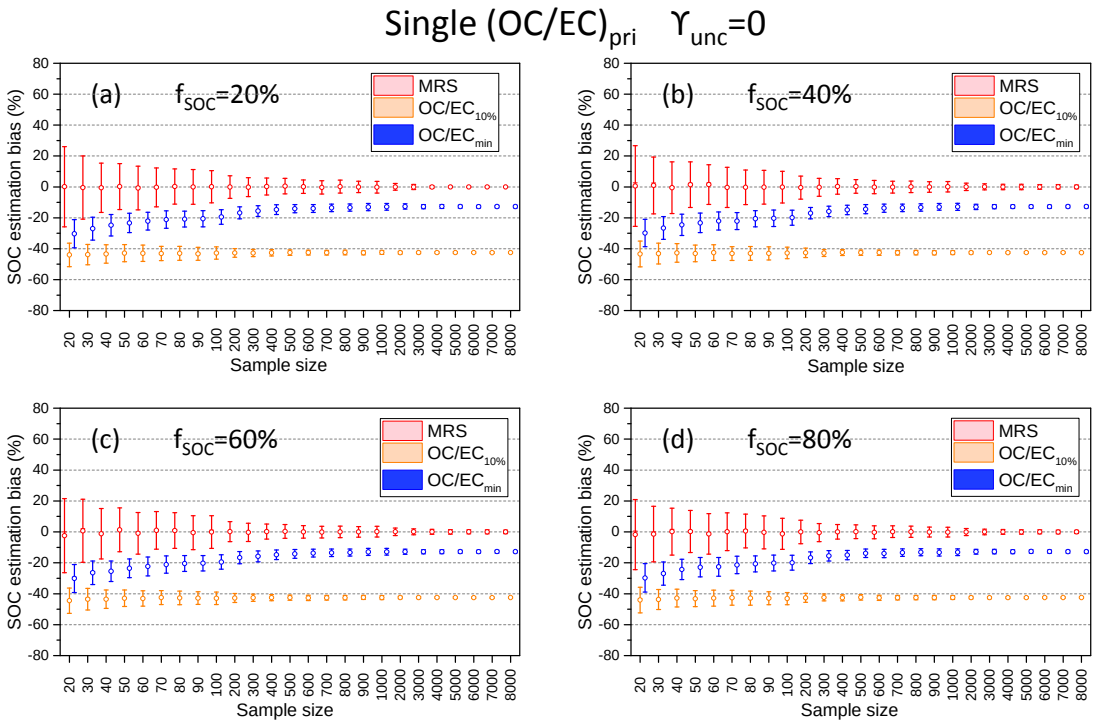


Figure S16. Bias of SOC determination as a function of sample size by different representation of $(OC/EC)_{pri}$, including: minimum R square method (MRS), $OC/EC_{10\%}$ and OC/EC_{min}. The empty circles represent mean of 500 repeat runs, the whiskers represent one standard deviation. Fixed input parameters: n = 20~8000, EC = 2±1 μgC m^{-3}, Υ_{unc}=0 and $(OC/EC)_{pri}$ = 0.5. f_{SOC} is varied, with (a) f_{SOC} = 20% SOC = 0.25±0.13 μgC m^{-3} (b) f_{SOC} = 40% SOC = 0.67±0.33 μgC m^{-3} (c) f_{SOC} = 60% SOC = 1.5±0. 75 μgC m^{-3} (d) f_{SOC} = 80% SOC = 4±2 μgC m^{-3}.

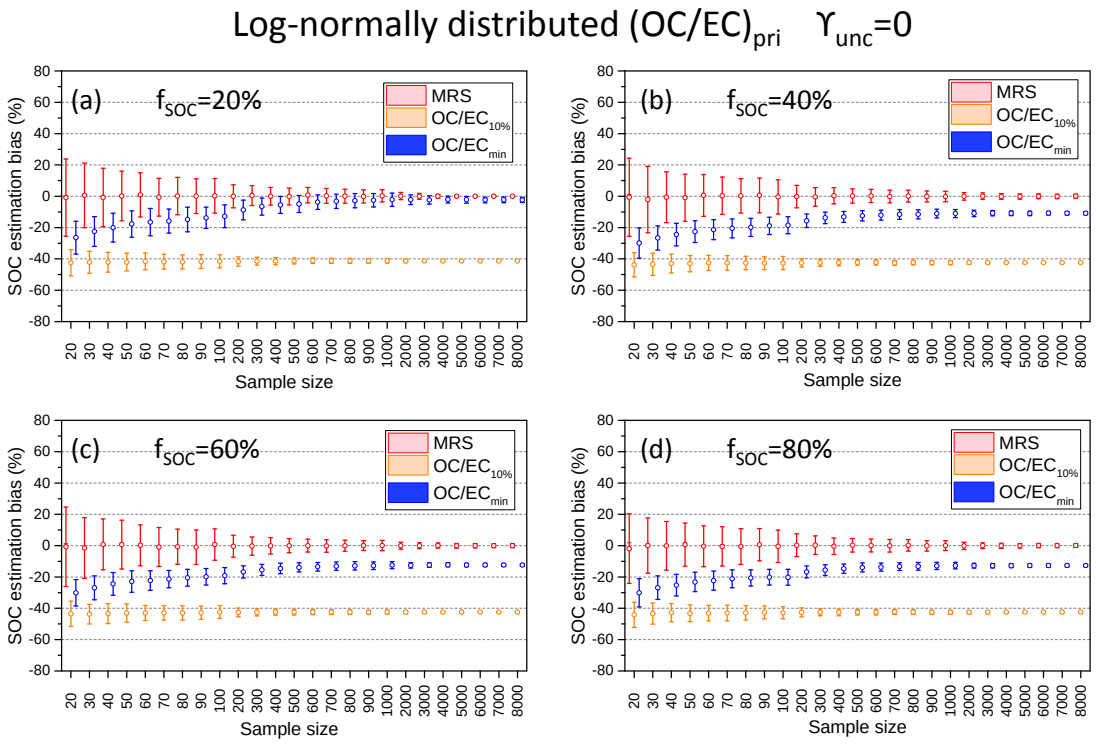


Figure S17. Bias of SOC determination as a function of sample size by different representation of $(OC/EC)_{pri}$, including: minimum R square method (MRS), $OC/EC_{10\%}$ and OC/EC_{min}. The empty circles represent mean of 500 repeat runs, the whiskers represent one standard deviation. Fixed input parameters: n = 20~8000, EC = 2±1 μgC m^{-3}, Υ_{unc}= 0 and $(OC/EC)_{pri}$ = 0.5±0.025. f_{SOC} is varied, with (a) f_{SOC} = 20% SOC = 0.25±0.13 μgC m^{-3} (b) f_{SOC} = 40% SOC = 0.67±0.33 μgC m^{-3} (c) f_{SOC} = 60% SOC = 1.5±0.75 μgC m^{-3} (d) f_{SOC} = 80% SOC = 4±2 μgC m^{-3}.

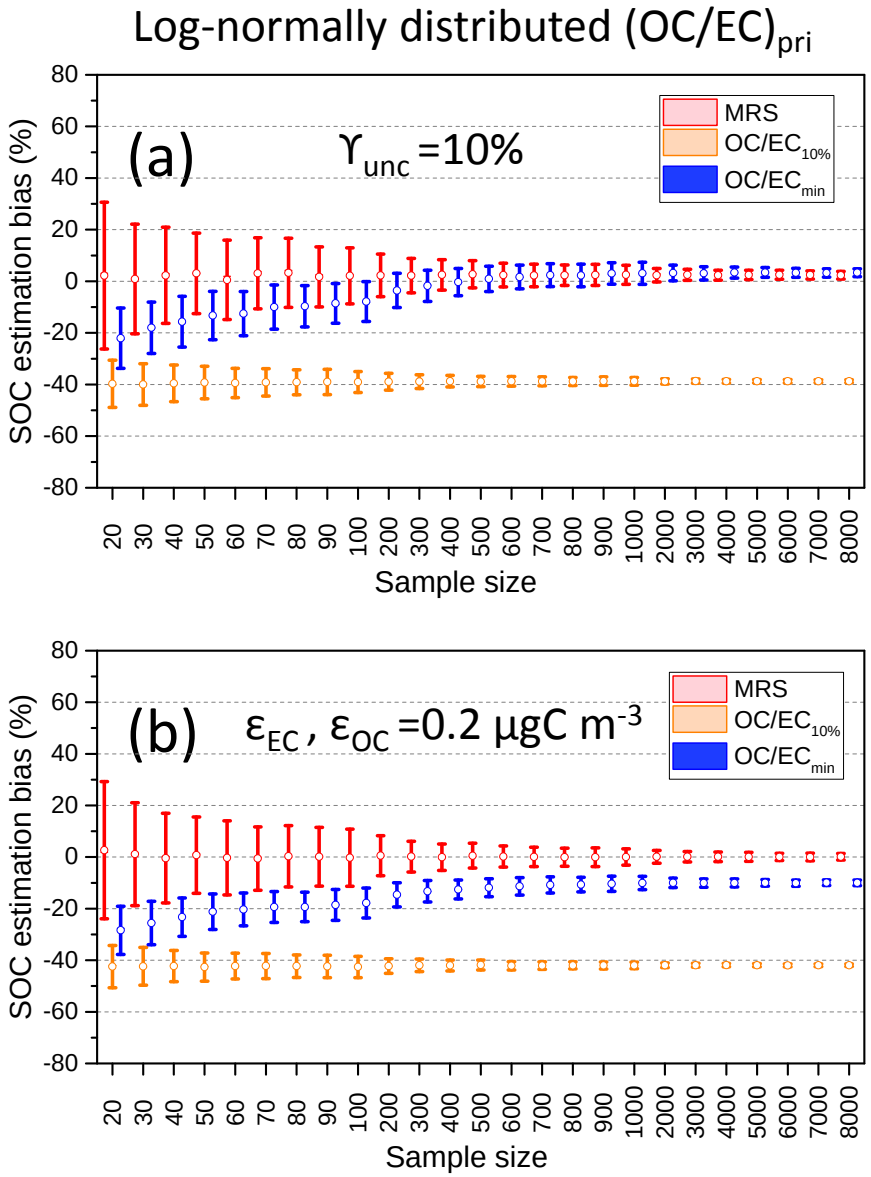


Figure S18. Bias of SOC determination as a function of sample size by different representation of $(OC/EC)_{pri}$, including: minimum R square method (MRS), $OC/EC_{10\%}$ and OC/EC_{min}. For each sample size, 500 repeat runs were conducted. The empty circles represent mean of 500 repeat runs, the whiskers represent one standard deviation. Fixed input parameters: n = 20~8000, EC = 8±4 μgC m^{-3}, f_{SOC} = 40% and $(OC/EC)_{pri}$ = 0.5±0.025. Two types of measurement uncertainties are considered (a) Υ_{unc} = 10% (b) ε_{EC}, ε_{OC} = 0.2 μgC m^{-3}.

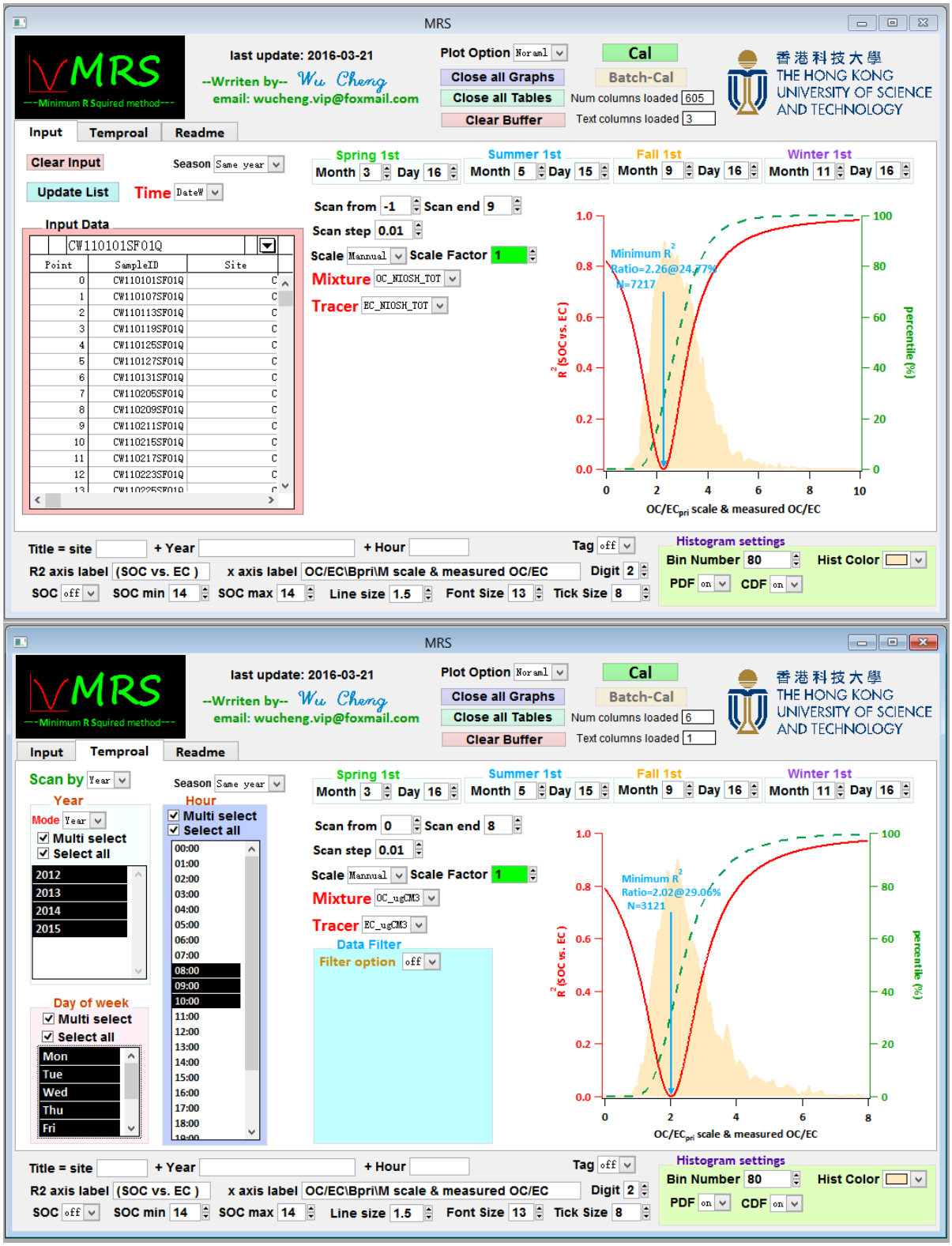
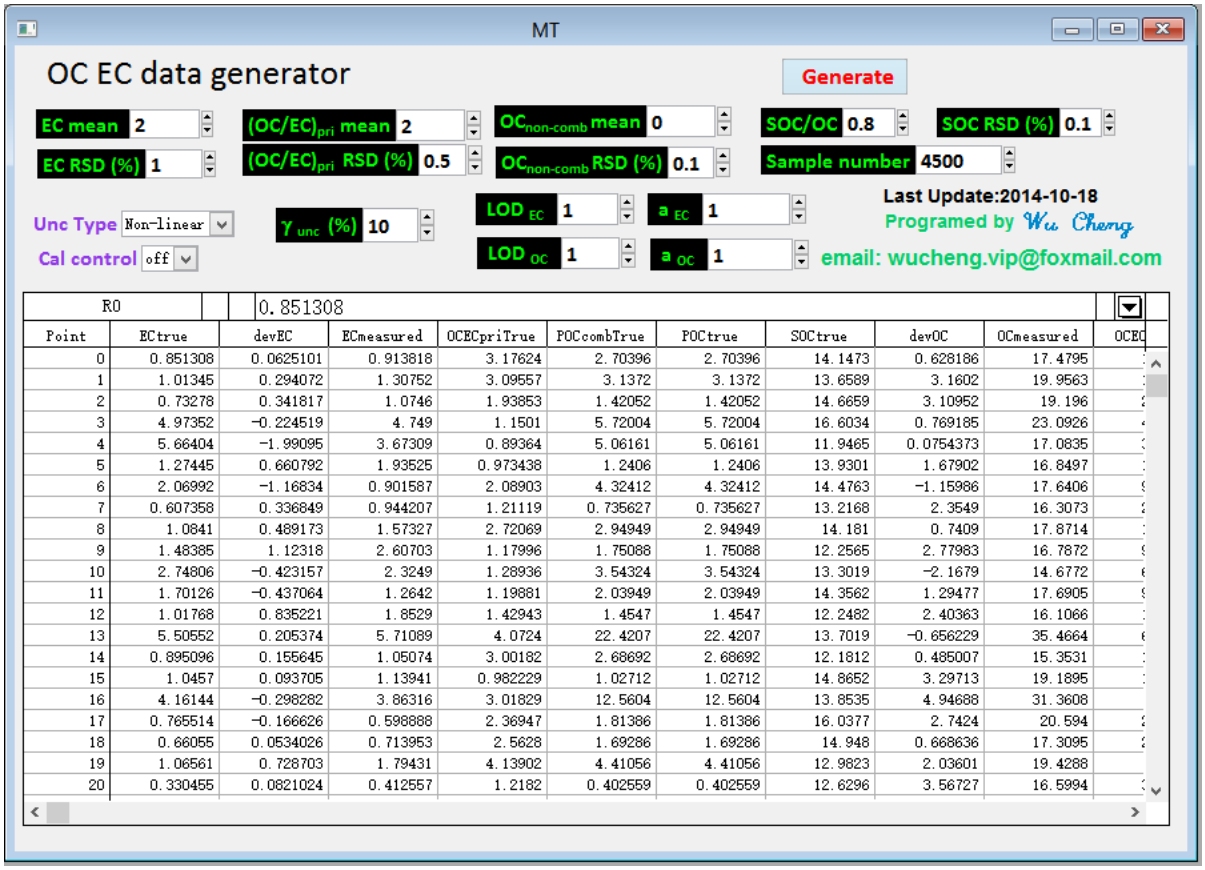


Figure S19. MRS Igor program. The calculation can be done on different time scale including hourly, day of week, monthly, seasonally and yearly. MRS can also be calculated on grouped data based on a text data column provided by the user.



Point	ECtrue	devEC	ECmeasured	OCECpriTrue	POCcombTrue	POCtrue	SOCtrue	devOC	OCmeasured	OCEC
0	0.851308	0.0625101	0.913818	3.17624	2.70396	2.70396	14.1473	0.628186	17.4795	
1	1.01345	0.294072	1.30752	3.09557	3.1372	3.1372	13.6589	3.1602	19.9563	
2	0.73278	0.341817	1.0746	1.93853	1.42052	1.42052	14.6659	3.10952	19.196	
3	4.97352	−0.224519	4.749	1.1501	5.72004	5.72004	16.6034	0.769185	23.0926	
4	5.66404	−1.99095	3.67309	0.89364	5.06161	5.06161	11.9465	0.0754373	17.0835	
5	1.27445	0.660792	1.93525	0.973438	1.2406	1.2406	13.9301	1.67902	16.8497	
6	2.06992	−1.16834	0.901587	2.08903	4.32412	4.32412	14.4763	−1.15986	17.6406	
7	0.607358	0.336849	0.944207	1.21119	0.735627	0.735627	13.2168	2.3549	16.3073	
8	1.0841	0.489173	1.57327	2.72069	2.94949	2.94949	14.181	0.7409	17.8714	
9	1.48385	1.12318	2.60703	1.17996	1.75088	1.75088	12.2565	2.77983	16.7872	
10	2.74806	−0.423157	2.3249	1.28936	3.54324	3.54324	13.3019	−2.1679	14.6772	
11	1.70126	−0.437064	1.2642	1.19881	2.03949	2.03949	14.3562	1.29477	17.6905	
12	1.01768	0.835221	1.8529	1.42943	1.4547	1.4547	12.2482	2.40363	16.1066	
13	5.50552	0.205374	5.71089	4.0724	22.4207	22.4207	13.7019	−0.656229	35.4664	
14	0.895096	0.155645	1.05074	3.00182	2.68692	2.68692	12.1812	0.485007	15.3531	
15	1.0457	0.093705	1.13941	0.982229	1.02712	1.02712	14.8652	3.29713	19.1895	
16	4.16144	−0.298282	3.86316	3.01829	12.5604	12.5604	13.8535	4.94688	31.3608	
17	0.765514	−0.166626	0.598888	2.36947	1.81386	1.81386	16.0377	2.7424	20.594	
18	0.66055	0.0534026	0.713953	2.5628	1.69286	1.69286	14.948	0.668636	17.3095	
19	1.06561	0.728703	1.79431	4.13902	4.41056	4.41056	12.9823	2.03601	19.4288	
20	0.330455	0.0821024	0.412557	1.2182	0.402559	0.402559	12.6296	3.56727	16.5994	

Figure S20. MT Igor program. OC and EC data following log-normal distribution can be generated for statistical study purpose (no time series information). User can define mean and RSD of EC, $(OC/EC)_{pri}$, SOC/OC ratio, measurement uncertainty, sample size, etc.