Electronic supplementary information

"Multiple sites ground-based evaluation of carbonaceous aerosol mass concentrations retrieved from the CAMS and MERRA-2 over the Indo-Gangetic Plain"

Ashish Soni [a], Anil Kumar Mandariya [a], Pradhi Rajeev [a], Saifi Izhar [a], Gyanesh Kumar Singh [a], Vikram Choudhary [a], Adnan Mateen Qadri [a], Aman Deep Gupta [a], Amit Kumar Singh [a] and Tarun Gupta [a,*]

[a] Department of Civil Engineering and APTL at Center for Environmental Science and Engineering (CESE), Indian Institute of Technology, Kanpur-208016, India

*Corresponding Author: Tarun Gupta, tarun@iitk.ac.in

List of Sections:

List of Tables:

List of Figures:

color. Fig S1 (a) and (b) represents relative error in BC mass concentrations; Fig S1 (c) and (d) represents relative error in OC mass concentrations

Fig S2. Relative error in CAMSRA and MERRA-2 simulated BC and OC mass concentrations at Kanpur. The CAMSRA data are shown with pink color, while MERRA-2 data are shown in Cyan color. Fig S2 (a) and (b) represents relative error in BC mass concentrations; Fig S2 (c) and (d) represents relative error in OC mass concentrations

Fig S3. Relative error in CAMSRA and MERRA-2 simulated BC and OC mass concentrations at Allahabad. The CAMSRA data are shown with pink color, while MERRA-2 data are shown in Cyan color. Fig S3 (a) and (b) represents relative error in BC mass concentrations; Fig S3 (c) and (d) represents relative error in OC mass concentrations

Fig S4. Relative error in CAMSRA and MERRA-2 simulated BC and OC mass concentrations at Lumbini. The CAMSRA data are shown with pink color, while MERRA-2 data are shown in Cyan color. Fig S4 (a) and (b) represents relative error in BC mass concentrations; Fig S4 (c) and (d) represents relative error in OC mass concentrations

Fig S5. Diurnal variation of ground-based measured BC and OC mass concentrations along with CAMSRA simulated mass concentrations of BC and OC over the Kanpur. Shaded area represents the standard deviations. Fig S5 (a) represents diurnal variations of BC, and Fig S5 (b) represents diurnal variations of OC

Fig S6. Diurnal variation of relative error associated with the MERRA-2 simulated mass concentrations of BC and OC over the Kanpur. Measurement dates are shown in the right side legends. Fig S6 (a) represents relative error of BC, and Fig S6 (b) represents relative error of OC

Section S1. Conversion of carbonaceous aerosol mass concentrations from PM_1 to $PM_{2.5}$

In the present study, BC or OC associated with $PM_{2.5}$ were compared with the CAMSRA and MERRA-2 reanalysis data of total BC or OC. Previous studies have reported that the mass median diameter of carbonaceous BC and OC falls under 0.3μm and 0.8μm, respectively.[1,2] Therefore, comparing these measured data from $PM_{2.5}$ with reanalysis data could be less biased. It is worth mentioning that some of the ground-based data at Kanpur (data of the year 2014) were measured from the PM_1 samples. We have extrapolate these carbonaceous aerosols data of PM_1 into $PM_{2.5}$ to make the ground-based data homogeneous. An earlier study by Gupta and Mandariya[3] reported that PM_1 and $PM_{2.5}$ ratio is equal to 0.85 at Kanpur. In the present study, similar scaling factor of 0.85 was utilized to convert carbonaceous aerosols associated with PM_1 to $PM_{2.5}$. As mentioned earlier, the mass median diameter of BC and OC falls under 0.8μm. Thus, the use of identical scaling factor of 0.85 to convert BC and OC mass concentrations from PM_1 to $PM_{2.5}$ could be less biased.

References:

1. Cheng Y, Li SM, Gordon M, Liu P. Size distribution and coating thickness of black carbon from the Canadian oil sands operations. Atmos Chem Phys. 2018;18(4):2653–67.

2. Wang N, Yu JZ. Size distributions of hydrophilic and hydrophobic fractions of water-soluble organic carbon in an urban atmosphere in Hong Kong. Atmos Environ [Internet]. 2017;166:110–9. Available from: http://dx.doi.org/10.1016/j.atmosenv.2017.07.009

3. Gupta T, Mandariya A. Sources of submicron aerosol during fog-dominated wintertime at Kanpur. Environ Sci Pollut Res. 2013;20(8):5615–29.

Section S2. Details about the reanalysis data performance statistics

Several statistical analysis were performed in the present study to get more insight into the association between ground-based carbonaceous aerosols with reanalysis data. Some of the basic statistics such as relative error, Mean Percentage Error (MPE), Root Mean Square Error (RMSE), and Pearson's correlation coefficient (R) are calculated using the below-mentioned equations:

$$Relative\ Error = \left(\frac{C_{e(i)} - C_{m(i)}}{C_{m(i)}}\right) \quad \text{Equation 1}$$

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{C_{e(i)} - C_{m(i)}}{C_{m(i)}}\right) \times 100 \quad \text{Equation 2}$$

$$RMSE = \left[\frac{\sum_{i=1}^{n}(C_{e(i)} - C_{m(i)})^2}{n}\right]^{\frac{1}{2}} \quad \text{Equation 3}$$

$$R = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(\frac{C_{e(i)} - \bar{C}_e}{\sigma_e}\right) \times \left(\frac{C_{m(i)} - \bar{C}_m}{\sigma_m}\right) \quad \text{Equation 4}$$

where $C_{e(i)}$ and $C_{m(i)}$, are i^{th} estimated and measured mass concentrations; $\bar{C}_e$ and $\bar{C}_m$ are average mass concentrations of estimated and measured mass concentrations; σ_e and σ_m are the standard deviation of estimated and measured mass concentrations; n is the total number of observations.

The relative error measures the error in reanalysis data concerning the actual measured value (true value). The MPE value provides average information about the reanalysis simulations and measured values. It can be both positive or negative, indicating the average amount of overestimation and underestimation of reanalysis data, respectively. The RMSE value provides information on the performance of reanalysis simulation. It is always a positive value, and for better performance of any simulations, low RMSE is desirable. The correlation coefficient (R) checks the linear relationship between measured and estimated or reanalysis mass concentrations.

It can be both positive or negative value between -1 and +1. For better performance of any simulations, the correlation coefficient should approach unity as close as possible.

Moreover, several other statistics were calculated to check the performance of CAMSRA and MERRA-2 reanalysis data. Details about the statistics are given below:

1. The Fraction of predictions within a factor or two (FAC2): It used to compare all the paired observations among estimated and measured data, it returns a value indicating fraction of paired observations which comes under the range given in the formula with the total number of paired observations.

   Formula: $$0.5 \leq \left[\frac{C_e}{C_m}\right] \leq 2$$ Equation 5

2. Mean bias (MB): It is used to find average bias between the estimated and measured data.

   Formula: $$MB = \frac{1}{n}\sum_{i=1}^{n} C_{e(i)} - C_{m(i)}$$ Equation 6

3. Mean Gross error (MGE): It is the average of absolute difference between paired observations i.e. between estimated and measured data points.

   Formula: $$MGE = \frac{1}{n}\sum_{i=1}^{n} |C_{e(i)} - C_{m(i)}|$$ Equation 7

4. Normalised Mean Bias (NMB): In this mean bias is normalised with the measured data.

   Formula: $$NMB = \frac{(\sum_{i=1}^{n} C_{e(i)} - C_{m(i)})}{(\sum_{i=1}^{n} C_{m(i)})}$$ Equation 8

5. Normalised Mean Gross Error (NMGE): It normalises the gross mean error with measured values.

$$NMGE = \frac{(\sum_{i=1}^{n} |C_{e(i)} - C_{m(i)}|)}{(\sum_{i=1}^{n} C_{m(i)})}$$

Formula: Equation 9

Table S1. Pearson correlation coefficient (R) between absolute relative error in reanalysis data of BC (OC) with meteorological data and chemical marker species normalized with carbonaceous aerosols.

Study sites		Temp. in °C	RH in %	Wind speed in m s^{-1}	SO_4^{2-}/BC (SO_4^{2-}/OC)	NO_3^-/BC (NO_3^-/OC)	NH_4^+/BC (NH_4^+/OC)	Cl (Cl
S1-Beas	CAMSRA	0.18 (-0.3)	-0.34 **(0.46*)**	0.29 (**-0.46***)	-0.27 (0.34)	-0.21 (0.08)	-0.16 (0.15)	-0 **(0.**
	MERRA-2	-0.07 (-0.23)	**-0.53*** (-0.19)	**0.37*** **(0.39*)**	**-0.57*** (-0.34)	0.07 (0.16)	-0.23 (-0.08)	-0 (-0
S2-Kanpur	CAMSRA	**-0.59*** **(-0.64*)**	**0.46*** **(0.57*)**	**-0.27*** **(-0.38*)**	**0.3*** (0.02)	**0.79*** **(0.75*)**	**0.67*** **(0.48*)**	**0.** **(0.**
	MERRA-2	**-0.58*** **(0.38*)**	**0.43*** **(-0.31*)**	-0.1 **(0.47*)**	**0.22*** **(-0.31*)**	**0.85*** **(-0.3*)**	**0.67*** **(-0.41*)**	**0.** **(-0**
S3-Allahabad	CAMSRA	0.11 (-0.01)	-0.29 (0.05)	-0.11 (0.16)	-0.33 **(0.56*)**	-0.26 (0.3)	-0.29 (0.38)	-0 (-0
	MERRA-2	0.27 (-0.22)	-0.25 (0.09)	-0.03 (0.12)	-0.36 **(-0.59*)**	**-0.69*** **(-0.64*)**	**-0.61*** **(-0.62*)**	-0. (-0
S4-Lumbini	CAMSRA	0.15 (0.08)	-0.16 (-0.29)	0.18 (0.07)	0.21 **(0.38*)**	-0.02 (-0.06)	0.03 (0.13)	-0 **(-0.**
	MERRA-2	0.09 (-0.1)	-0.17 (0.13)	0.17 (0.09)	0.14 (-0.03)	-0.03 (-0.03)	0.02 (-0.04)	-0 **(0.**

** represents the correlation is statistical significant at 95% confidence interval (p<0.05); POC: Primary Org Carbon; WSOC: Water Soluble Organic Carbon; ALWC: Aerosol Liquid Water Content*

Table S2. Ratio of HCl and OC or EC emitted from the different emission sources given by Andreae (2019).

Types of sources (Burning of)	HCl/OC	HCl/BC
Agricultural residual (open)	0.04	0.43
Biofuels (without dung)	0.02	0.09
Dung	0.004	0.12
Charcoal	0.005	0.41
Garbage	0.51	2

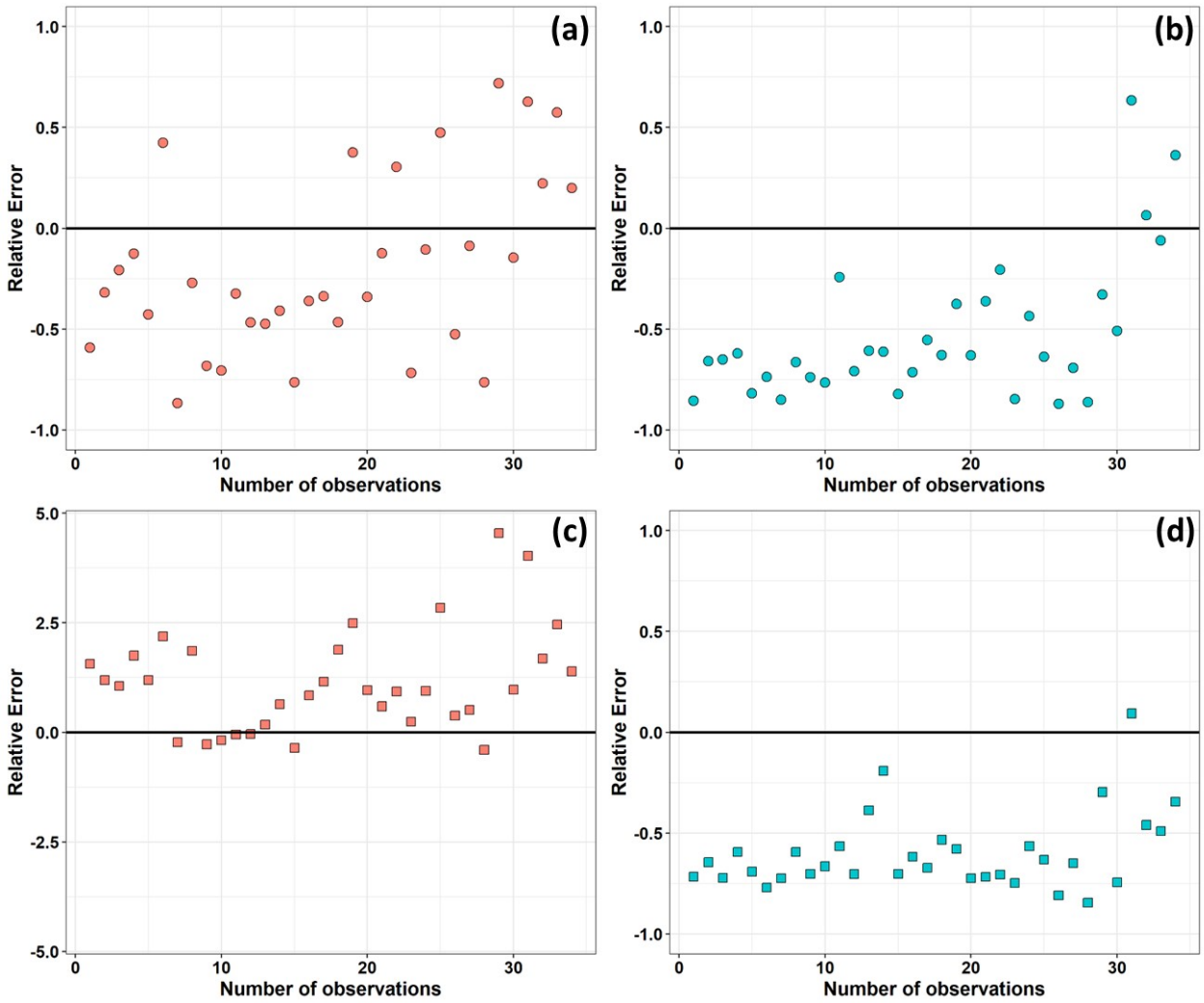

Fig S1. Relative error in CAMSRA and MERRA-2 simulated BC and OC mass concentrations at Beas. The CAMSRA data are shown with pink color, while MERRA-2 data are shown in Cyan color. Fig S1 (a) and (b) represents relative error in BC mass concentrations; Fig S1 (c) and (d) represents relative error in OC mass concentrations

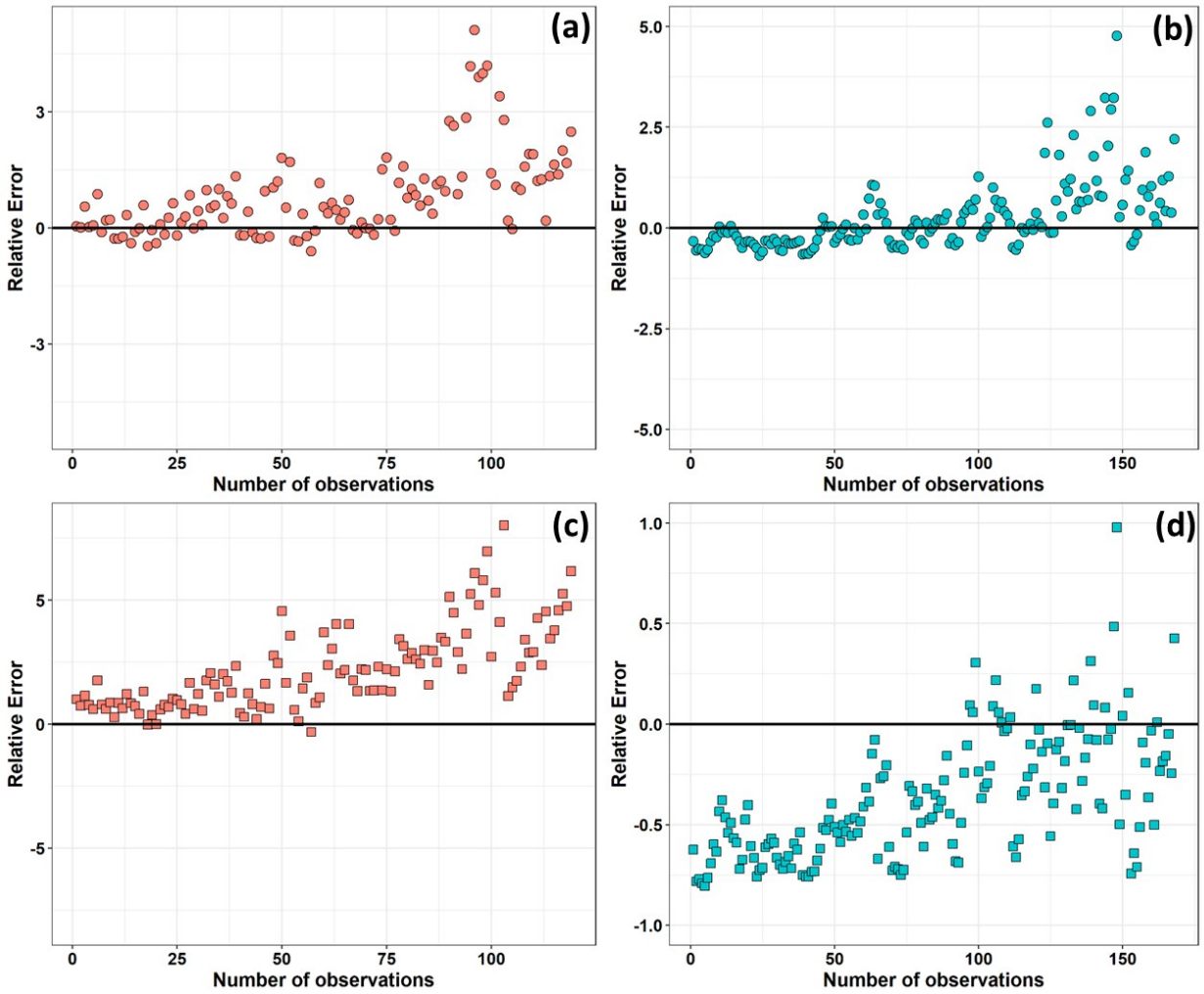

Fig S2. Relative error in CAMSRA and MERRA-2 simulated BC and OC mass concentrations at Kanpur. The CAMSRA data are shown with pink color, while MERRA-2 data are shown in Cyan color. Fig S2 (a) and (b) represents relative error in BC mass concentrations; Fig S2 (c) and (d) represents relative error in OC mass concentrations

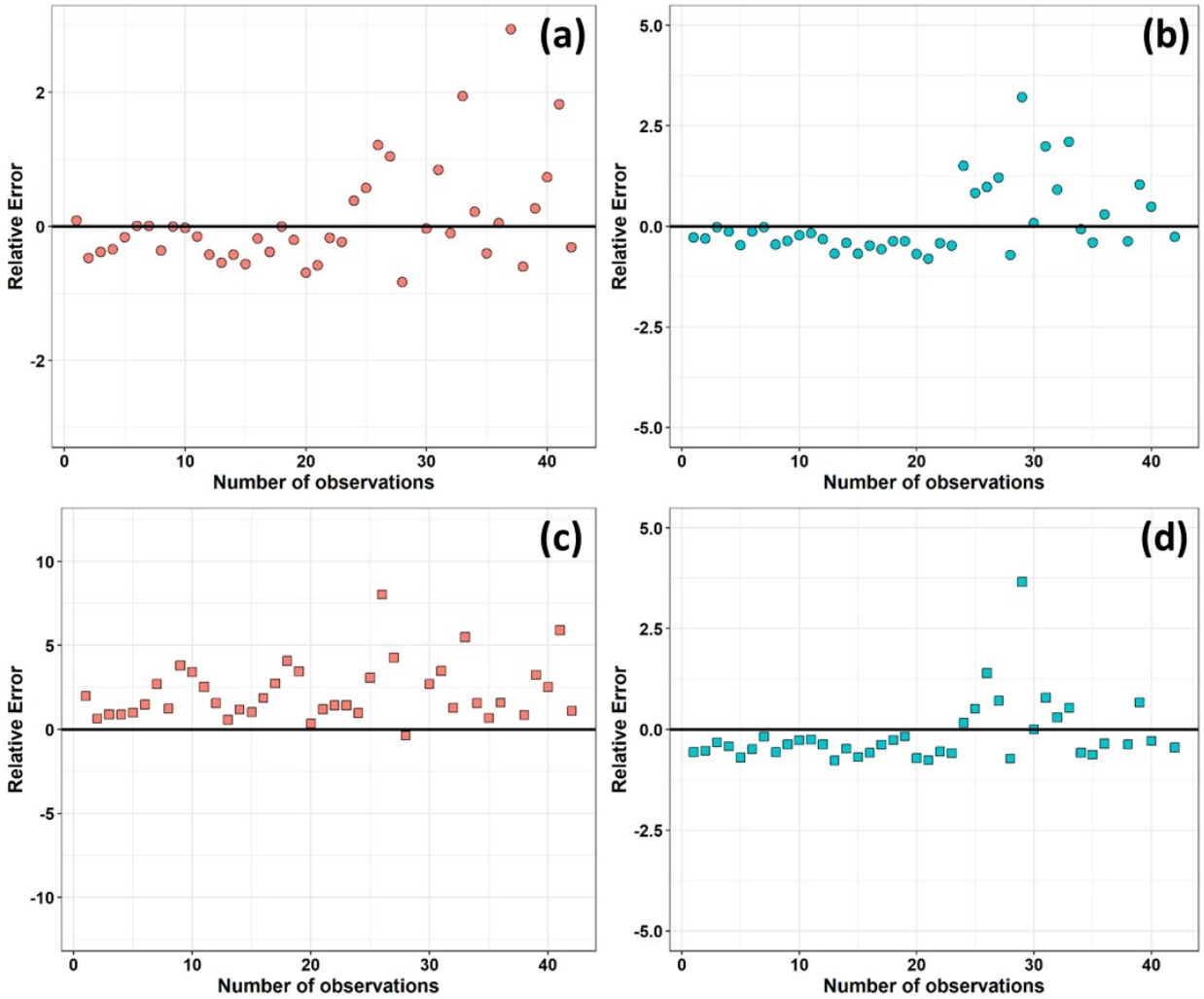

Fig S3. Relative error in CAMSRA and MERRA-2 simulated BC and OC mass concentrations at Allahabad. The CAMSRA data are shown with pink color, while MERRA-2 data are shown in Cyan color. Fig S3 (a) and (b) represents relative error in BC mass concentrations; Fig S3 (c) and (d) represents relative error in OC mass concentrations

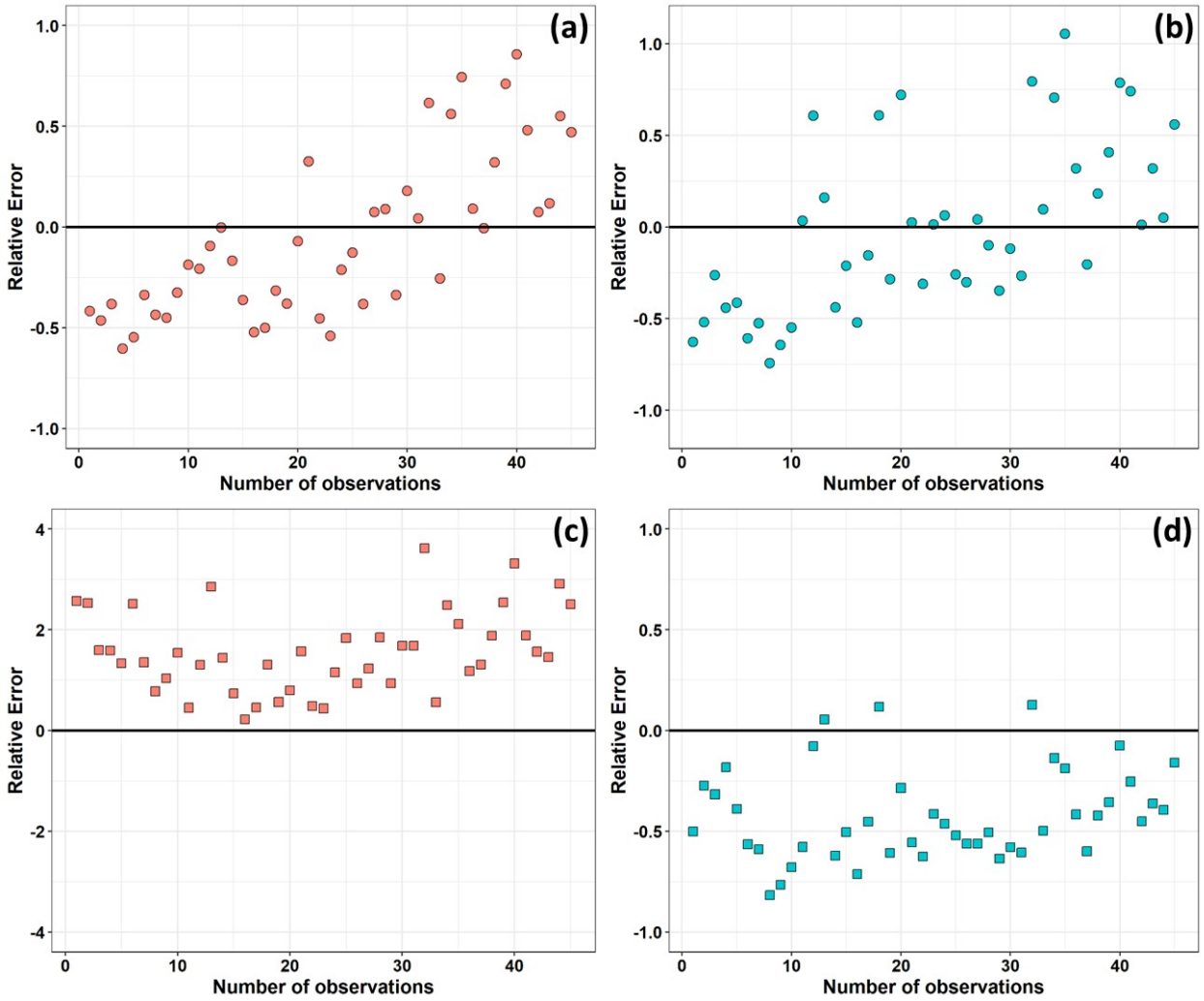

Fig S4. Relative error in CAMSRA and MERRA-2 simulated BC and OC mass concentrations at Lumbini. The CAMSRA data are shown with pink color, while MERRA-2 data are shown in Cyan color. Fig S4 (a) and (b) represents relative error in BC mass concentrations; Fig S4 (c) and (d) represents relative error in OC mass concentrations

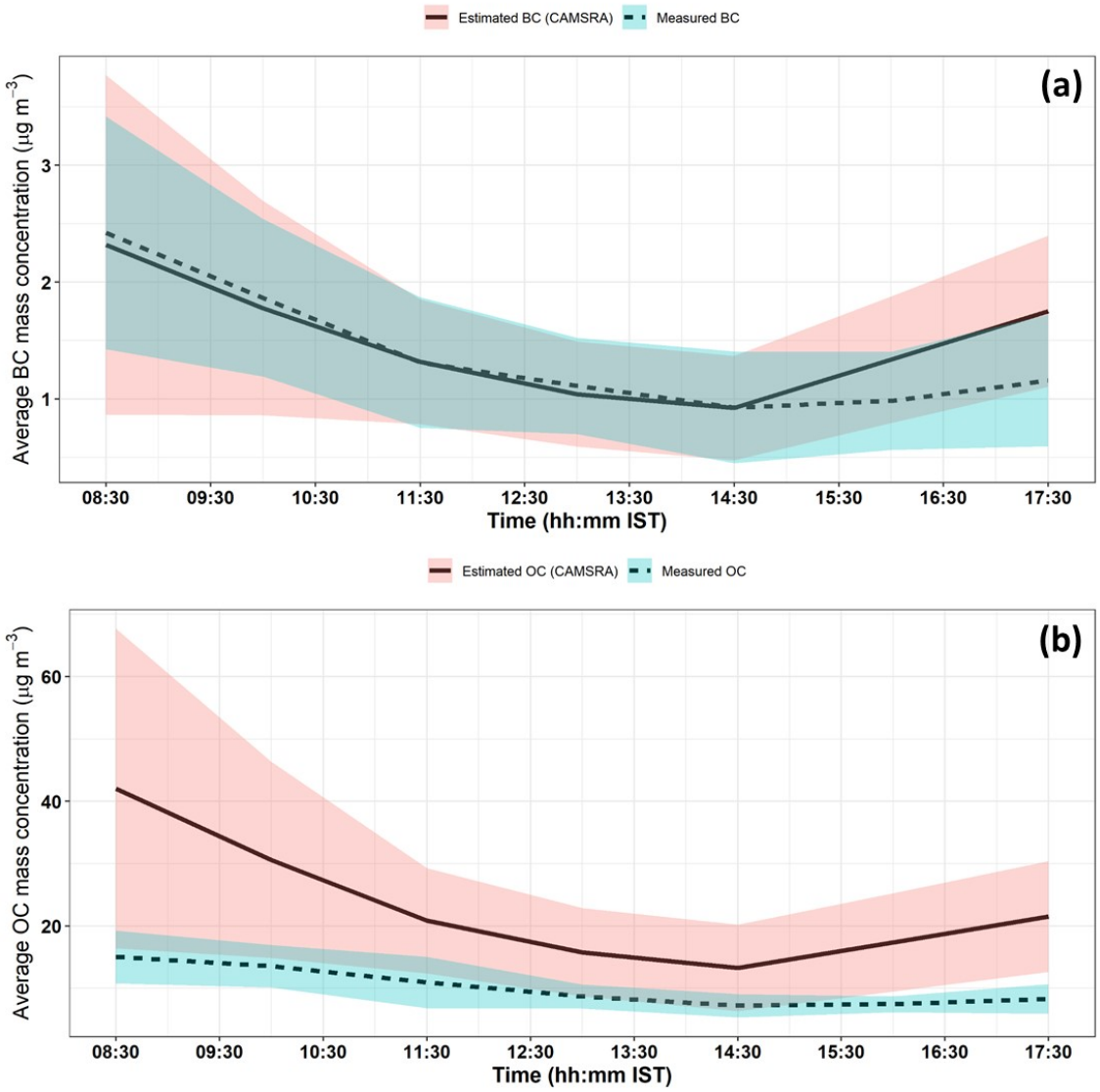

Fig S5. Diurnal variation of ground-based measured BC and OC mass concentrations along with CAMSRA simulated mass concentrations of BC and OC over the Kanpur. Shaded area represents the standard deviations. Fig S5 (a) represents diurnal variations of BC, and Fig S5 (b) represents diurnal variations of OC

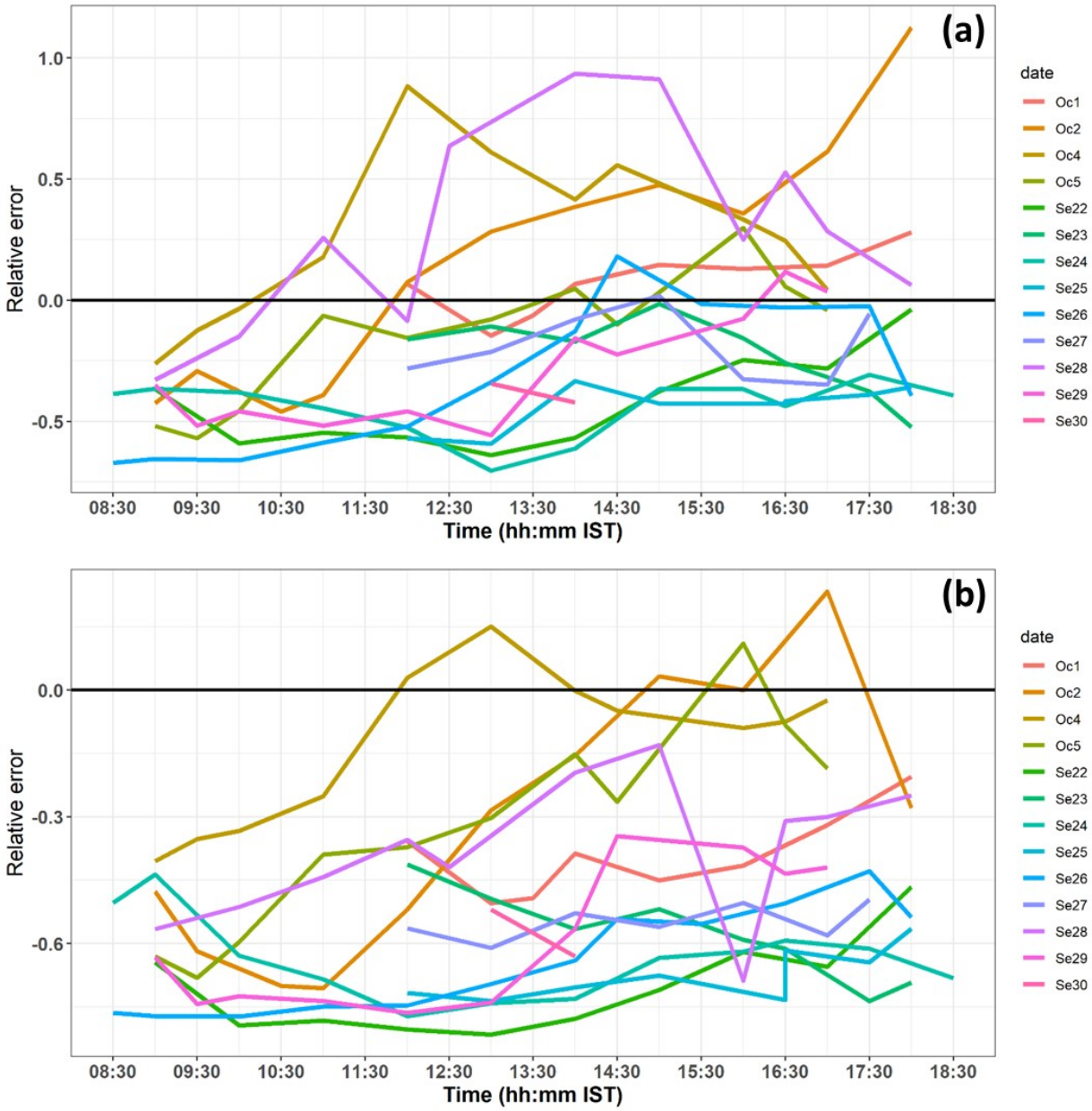

Fig S6. Diurnal variation of relative error associated with the MERRA-2 simulated mass concentrations of BC and OC over the Kanpur. Measurement dates are shown in the right side legends. Fig S6 (a) represents relative error of BC, and Fig S6 (b) represents relative error of OC