



*Supplement of*

## **Simulating organic aerosol in Delhi with WRF-Chem using the volatility-basis-set approach: exploring model uncertainty with a Gaussian process emulator**

**Ernesto Reyes-Villegas et al.**

*Correspondence to:* Gordon McFiggans ([g.mcfiggans@manchester.ac.uk](mailto:g.mcfiggans@manchester.ac.uk))

The copyright of individual parts of the supplement might differ from the article licence.

## S1. WRF-Chem setup.

Table S1 is based on Table 2 of Tsimpidi et al (2010), showing the aerosol yields for high- and low- $\text{NO}_x$  parameterisations for the SOA products we have included in this model, along with the CRI-v2R5 precursors we have used in this study. For full details of the scheme please refer to Tsimpidi et al (2010).

Table S1: SOA yield scenarios using a four-product basis set with saturations concentrations of 1, 10, 100, and  $1000 \mu\text{g m}^{-3}$  at 298 K.

V-SOA precursors	CRI-v2R5 precursor(s)	Aerosol Yield High $\text{NO}_x$ Parameterization				Aerosol Yield Low- $\text{NO}_x$ Parameterization			
		1	10	100	1000	1	10	100	1000
ARO1	TOLUENE, BENZENE	0.003	0.165	0.300	0.435	0.075	0.225	0.375	0.525
ARO2	OXYL	0.002	0.195	0.300	0.435	0.075	0.300	0.375	0.525
ISOP	ISOPRENE	0.001	0.023	0.015	0.000	0.009	0.030	0.015	0.000
TERP	APINENE	0.012	0.122	0.201	0.500	0.107	0.092	0.359	0.600

Table S2 is the mapping of WACCM6 chemical species to CRI-v2R5 & MOSAIC chemical species. The mapping of the inorganic aerosol components, and aerosol number, is taken from the mapping advice provided by Emmons, Pfister and Hodzic (2019), and not repeated here. The mapping of OA to VBS compounds has been estimated from our preliminary modelling studies, and is designed to provide ‘aged’ VBS aerosol at the boundaries of the domain. VBS volatility bins 1-3 ( $\log_{10}(C^*)$  values of -2 to 0) only are used, and a ‘non-oxygen’ to oxygen ratio of 1.925 has been adopted (estimated from our preliminary modelling studies). We add the VBS compounds as gases, and allow the model to dynamically partition these to the condensed-phase as needed. The WRF-Chem variable names for each of these has been given in the Table S2.

Table S2: Species mapping between WACCM6 and CRI-v2R5 chemical schemes. The scaling factor has been included where this is not 1. The scaling factor for the gas-phase VBS compounds includes the  $\text{kg kg}^{-1}$  to ppm conversion factor of  $1.1588 \times 10^5$  (calculated using the same molar mass of  $250 \text{ g mol}^{-1}$  that is used with WRF-Chem for these compounds).

WACCM6	CRI-v2R5	Scaling Factor	WACCM6	CRI-v2R5	Scaling Factor
O3	O3		C2H6	C2H6	
SO2	SO2		C3H6	C3H6	
NO	NO		C3H8	C3H8	
NO2	NO2		CH3COCH3	KET	
NH3	NH3		CH3OH	CH3OH	
N2O5	N2O5		CH3OOH	CH3OOH	
PAN	PAN		CH4	CH4	
MPAN	MPAN		CRESOL + XYLENES	OXYL	
NOA	NOA		DMS	DMS	
HNO3	HNO3		H2O2	H2O2	
HO2NO2	HNO4		MEK	MEK	
CO	CO		TOLUENE	TOLUENE	
BENZENE	BENZENE		POM_A1 + SOA1_A2 + SOA1_A1 + SOA2_A2 + SOA2_A1 + SOA3_A2 + SOA3_A1 + SOA4_A2 + SOA4_A1 + SOA5_A2 + SOA5_A1	PCG1_F_C PCG1_F_O PCG2_F_C PCG2_F_O PCG3_F_C	$4.6122 \times 10^4$ $2.3962 \times 10^4$ $3.3471 \times 10^4$ $1.7389 \times 10^4$ $4.8044 \times 10^2$
BIGALK	NC4H10	1.25		PCG3_F_O	$2.4960 \times 10^2$
BIGENE	TBUT2ENE				
C2H2	C2H2				
C2H4	C2H4				
C2H5OH	C2H5OH				

Table S3 shows an example of the parameters used to control the VBS scheme when building the model setup for anthropogenic and biomass burning sources. 111 namelist.input files were designed using the ranges of the 10 parameters in table 2.

Table S3: Example of a namelist.input file with parameters to control the VBS scheme.

Parameter	Value
Biomass Burning _VBS_ scaling	1
Anthropogenic VBS scaling	1
Biomass Burning VBS ageing rate	1.00E-13
Anthropogenic VBS ageing rate	1.00E-13
Biomass Burning VBS Oxidation rate	0.075
Anthropogenic _VBS_ Oxidation rate	0.075
Biomass Burning VBS FRAC_1	0
Biomass Burning VBS FRAC_2	0.12
Biomass Burning VBS FRAC_3	0.24
Biomass Burning VBS FRAC_4	0.24
Biomass Burning VBS FRAC_5	0.21
Biomass Burning VBS FRAC_6	0.13
Biomass Burning VBS FRAC_7	0.04
Biomass Burning VBS FRAC_8	0
Biomass Burning VBS FRAC_9	0
Anthropogenic VBS FRAC_1	0
Anthropogenic VBS FRAC_2	0.12
Anthropogenic VBS FRAC_3	0.24
Anthropogenic VBS FRAC_4	0.24
Anthropogenic VBS FRAC_5	0.21
Anthropogenic VBS FRAC_6	0.13
Anthropogenic VBS FRAC_7	0.04
Anthropogenic VBS FRAC_8	0
Anthropogenic VBS FRAC_9	0

**Ageing rate** is the reaction rate for all VBS reactions in that scheme (in  $\text{cm}^3 \text{ molec}^{-1} \text{ s}^{-1}$ )

**Oxidation rate** is the fractional increase in oxidation of the VBS compounds per reaction step

**FRAC[1-9]** is the multiplier from the POA mass in the emission database, to give the emitted mass of VBS component in that volatility bin. Volatility bin 1 has a  $\log_{10}(C^*)$  value of -2, and for bins 2-8 this increases decadally (as illustrated in Figure S1), to a  $\log_{10}(C^*)$  value of 6 for bin 9.

**scaling** is a scaling factor applied to all **FRAC[1-9]** values for that scheme, usually with the aim of ensuring that the condensed VBS mass at time of emission is roughly equivalent to the involatile POA mass in emission database used.

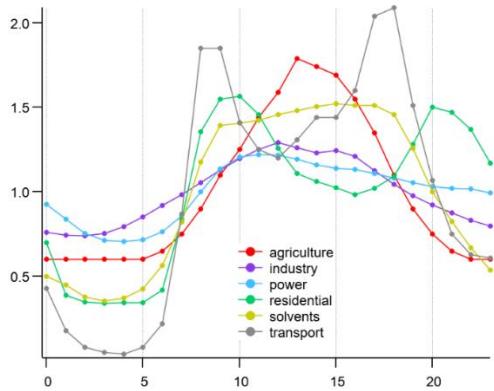


Figure S1 Diurnal fraction of seven activities used in WRF-Chem. Taken from (Olivier et al., 2003)

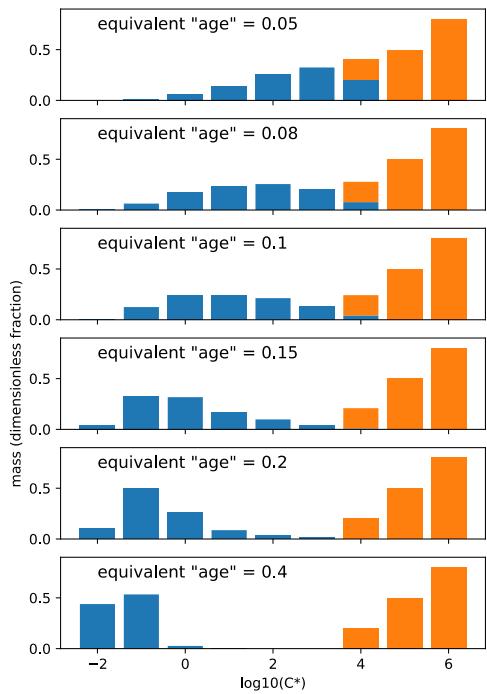


Figure S2 Example volatility distributions for emissions. Orange bars indicate IVOC emissions (with a fixed distribution). Blue bars indicate SVOC emissions, with volatility distributions calculated using the “equivalent age” calculation described in Section 2.3. For this plot both the IVOC and SVOC scaling factors are set to 1.

## S2. Observations

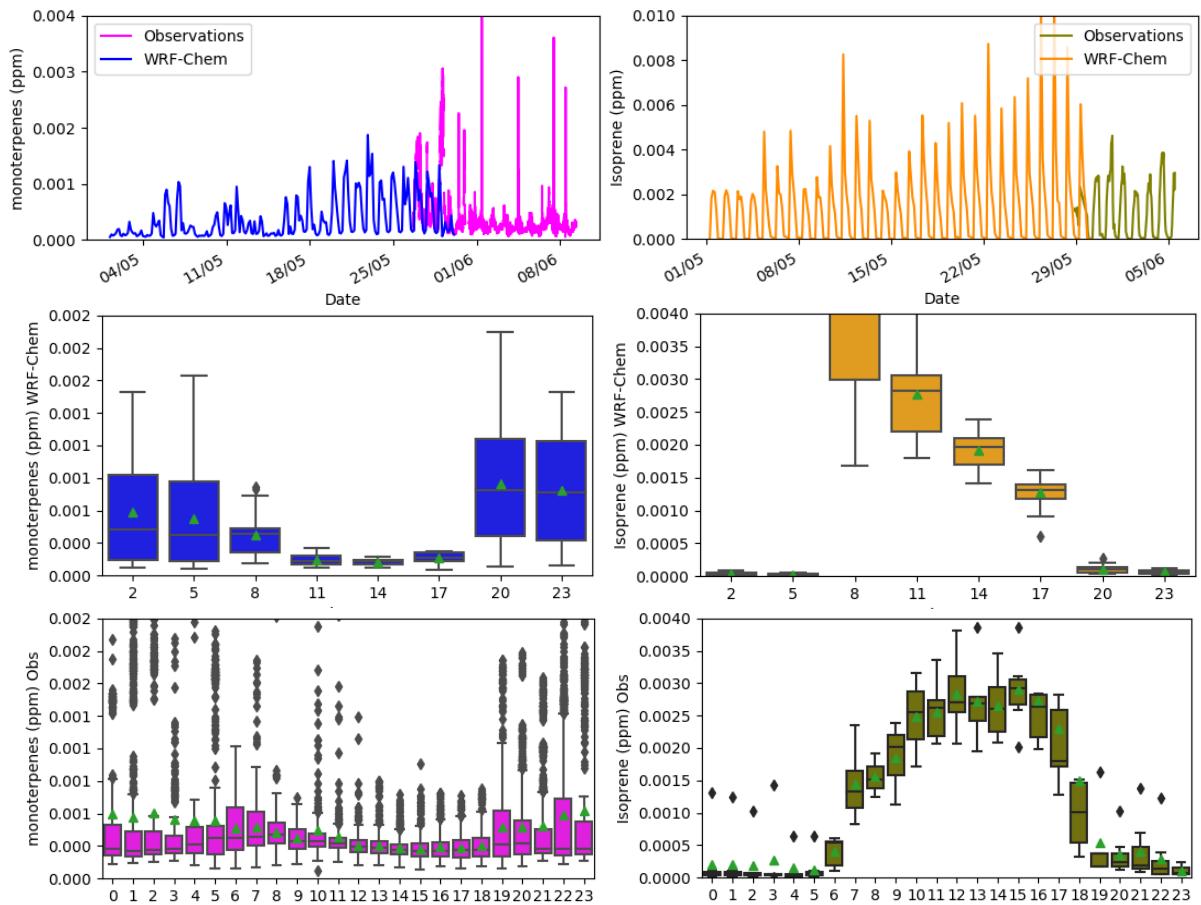


Figure S3. Monoterpenes (left) and isoprene (right) time series and diurnal cycles of WRF-Chem outputs and observations. Notice the different start-end times of model and observations. Triangles highlight the mean.

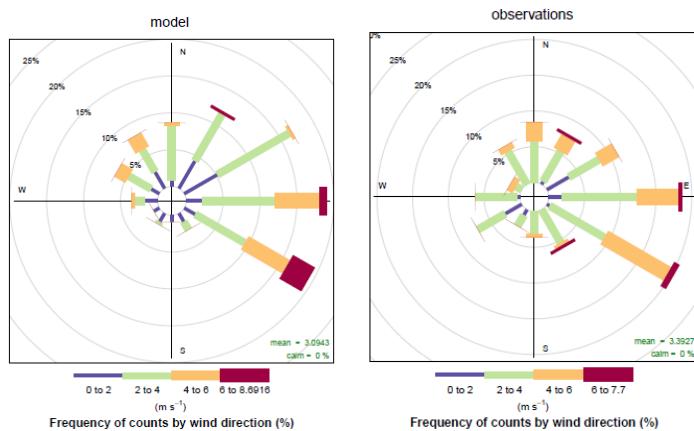
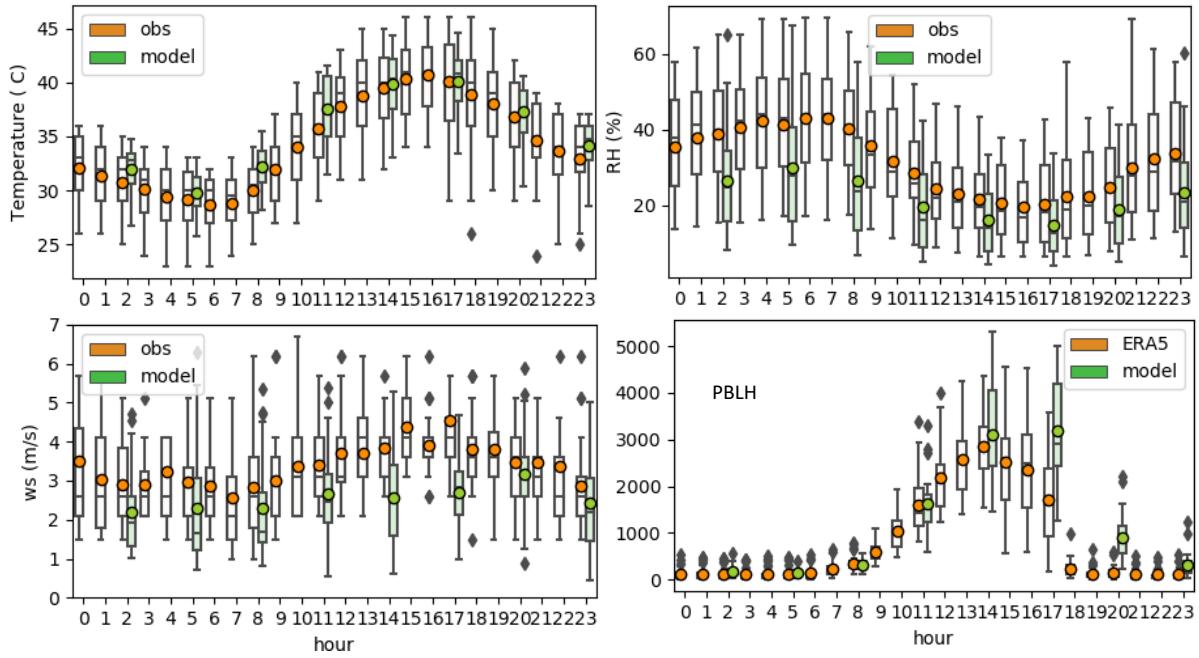


Figure S4. Wind roses and diurnal cycles of temperature, RH, ws and PBLH. May – 2018. Circles highlight the mean.

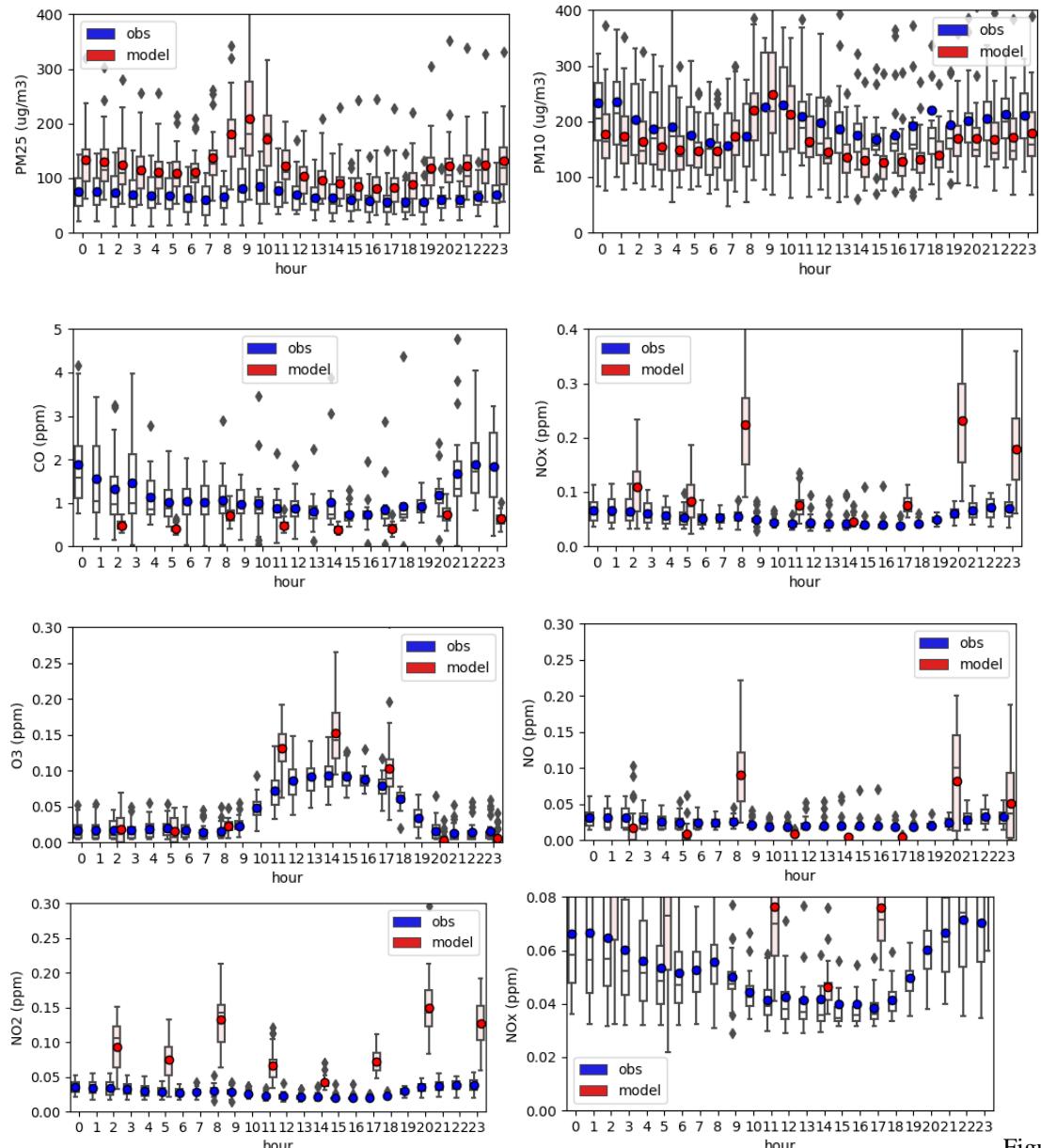


Figure S5.

Diurnal cycles of particulate matter and gaseous species. Observations and model outputs. May 2018. The circles highlight the mean.

### S3. WRF-Chem model evaluation

The parameters used for model evaluation were calculated with the OpenAir package (Carslaw and Ropkins, 2012). The following equations were extracted from the OpenAir manual, where  $O_i$  represents the  $i^{th}$  observed value and  $M_i$  represents the  $i^{th}$  modelled value for a total of n observations.

#### Fraction of predictions within a factor of two, FAC2

FAC2 is the fraction of modelled values within a factor of two of observations, which satisfy:

$$0.5 \leq \frac{M_i}{O_i} \leq 2.0$$

#### Mean bias (MB).

MB gives an indication of the mean over or underestimate of predictions; it has the same units as the quantities being considered.

$$MB = \frac{1}{n} \sum_{i=1}^n M_i - O_i$$

#### Index of agreement (IOA).

The IOA is commonly used in model evaluation (Willmott et al., 2012), ranging between -1 and +1, with values close to +1 representing a better model performance. An IOA of 0.5 indicates that the sum of the error magnitudes is one-half of the sum of the observed-deviation magnitudes. IOA, with  $c = 2$ , is defined as:

$$IOA = 1.0 - \frac{\sum_{i=1}^n |M_i - O_i|}{c \sum_{i=1}^n |O_i - \bar{O}|}, \text{ when}$$

$$\sum_{i=1}^n |M_i - O_i| \leq c \sum_{i=1}^n |O_i - \bar{O}|$$

$$IOA = \frac{c \sum_{i=1}^n |O_i - \bar{O}|}{\sum_{i=1}^n |M_i - O_i|} - 1.0, \text{ when}$$

$$\sum_{i=1}^n |M_i - O_i| > c \sum_{i=1}^n |O_i - \bar{O}|$$



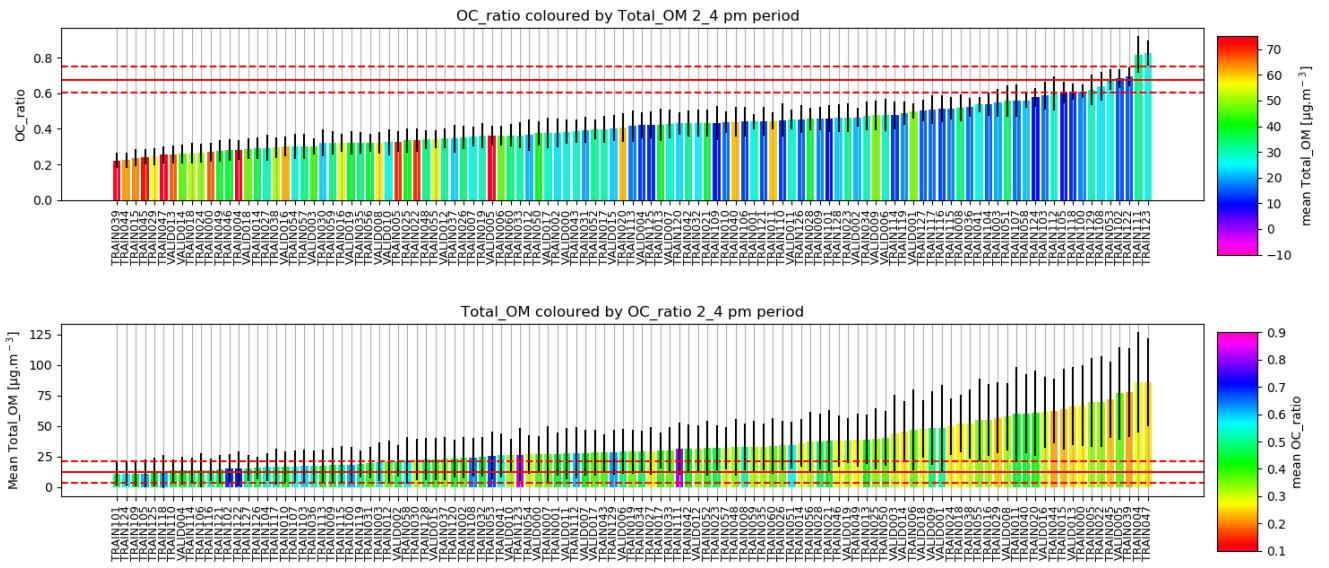


Figure S6. Analysis of the 111 model runs for the 2-4 pm period. Mean OC\_ratio coloured by mean Total\_OM (top) and mean Total\_OM coloured by mean OC\_ratio (bottom). The red line highlights the mean and SD of AMS observations (O:C top and OA bottom). The mean AMS values are O:C = 0.67 and OA = 12.20  $\mu\text{g.m}^{-3}$ .

Table S5 Evaluation of the ensemble of 111 model runs ordered from high to low FAC2 values with O:C and OA for the 2-4 pm period.

2-4pm_O:C						2-4pm_OA					
model	FAC2	MB	IOA	model	FAC2	MB	IOA	model	FAC2	MB	IOA
TRAIN036	1.00	0.03	0.51	TRAIN050	0.86	-0.13	0.36	TRAIN127	0.73	4.37	0.44
TRAIN114	1.00	0.04	0.45	TRAIN028	0.85	-0.10	0.45	VALID004	0.72	3.30	0.51
TRAIN127	0.99	0.02	0.51	TRAIN019	0.85	-0.12	0.39	TRAIN121	0.72	1.02	0.48
TRAIN101	0.99	0.02	0.48	TRAIN011	0.85	-0.13	0.37	TRAIN126	0.72	4.35	0.43
TRAIN117	0.99	0.04	0.48	TRAIN105	0.85	0.15	0.25	TRAIN110	0.70	2.03	0.53
TRAIN115	0.99	0.05	0.46	TRAIN040	0.84	-0.08	0.46	TRAIN125	0.70	0.27	0.50
TRAIN121	0.98	0.00	0.52	VALID010	0.84	-0.14	0.34	TRAIN124	0.38	-3.99	0.40
TRAIN106	0.98	0.00	0.52	TRAIN118	0.84	0.14	0.29	TRAIN010	0.70	6.93	0.49
TRAIN120	0.98	-0.04	0.52	TRAIN060	0.83	-0.14	0.33	TRAIN009	0.70	8.94	0.43
TRAIN119	0.98	0.01	0.51	TRAIN007	0.83	-0.16	0.25	TRAIN103	0.70	3.13	0.42
TRAIN109	0.98	0.02	0.50	TRAIN033	0.81	-0.13	0.37	TRAIN113	0.69	3.90	0.45
TRAIN126	0.98	0.01	0.50	TRAIN035	0.81	-0.16	0.27	TRAIN036	0.69	5.13	0.40
TRAIN058	0.98	0.05	0.48	TRAIN054	0.81	-0.16	0.27	TRAIN007	0.37	27.76	-0.22
TRAIN104	0.98	0.03	0.47	TRAIN048	0.81	-0.17	0.24	TRAIN031	0.69	9.43	0.40
TRAIN003	0.98	0.04	0.46	TRAIN053	0.81	0.18	0.20	TRAIN117	0.68	1.27	0.47
TRAIN034	0.96	-0.02	0.54	VALID008	0.81	-0.18	0.19	TRAIN104	0.68	5.50	0.47
TRAIN041	0.96	-0.03	0.52	TRAIN108	0.81	0.16	0.17	TRAIN116	0.68	2.09	0.43
TRAIN128	0.96	-0.02	0.52	TRAIN027	0.81	-0.15	0.30	TRAIN115	0.68	3.27	0.39
VALID006	0.96	-0.01	0.51	TRAIN026	0.80	-0.15	0.31	TRAIN015	0.68	3.27	0.39
VALID002	0.96	-0.04	0.51	TRAIN030	0.80	-0.15	0.30	TRAIN107	0.68	3.14	0.38
TRAIN008	0.96	-0.01	0.49	VALID012	0.80	-0.16	0.27	TRAIN048	0.32	31.49	-0.28
TRAIN125	0.96	-0.02	0.49	TRAIN025	0.80	-0.16	0.27	TRAIN102	0.67	10.89	0.39
TRAIN110	0.96	0.02	0.47	VALID019	0.80	-0.17	0.24	VALID011	0.35	29.12	-0.23
TRAIN009	0.95	-0.04	0.52	TRAIN059	0.80	-0.17	0.22	TRAIN119	0.67	7.12	0.35
TRAIN031	0.95	-0.07	0.51	TRAIN006	0.80	-0.17	0.21	TRAIN058	0.67	8.52	0.33
TRAIN051	0.95	0.02	0.49	VALID005	0.80	-0.19	0.17	VALID002	0.65	11.31	0.36
TRAIN107	0.95	0.09	0.36	TRAIN014	0.80	-0.19	0.17	TRAIN106	0.64	0.92	0.44
TRAIN112	0.95	0.10	0.31	TRAIN055	0.80	-0.19	0.15	TRAIN120	0.63	11.83	0.33
TRAIN042	0.94	-0.06	0.53	TRAIN102	0.80	0.20	0.08	TRAIN037	0.63	12.46	0.32
TRAIN129	0.94	0.11	0.30	TRAIN057	0.79	-0.17	0.22	TRAIN003	0.63	9.95	0.31
TRAIN103	0.94	0.11	0.29	TRAIN022	0.79	-0.18	0.19	TRAIN053	0.63	10.41	0.28
VALID007	0.93	-0.06	0.53	TRAIN056	0.79	-0.19	0.18	TRAIN129	0.63	14.04	0.25
TRAIN113	0.93	-0.04	0.52	TRAIN122	0.79	0.21	0.08	TRAIN108	0.62	11.43	0.28
TRAIN010	0.93	-0.07	0.50	TRAIN005	0.78	-0.20	0.12	TRAIN123	0.62	10.77	0.26
VALID004	0.93	-0.04	0.49	TRAIN038	0.78	-0.21	0.09	TRAIN128	0.59	11.86	0.26
VALID015	0.93	-0.09	0.49	VALID016	0.74	-0.21	0.08	TRAIN002	0.58	13.48	0.31
TRAIN021	0.93	-0.07	0.47	TRAIN016	0.74	-0.21	0.07	TRAIN112	0.58	15.74	0.14
TRAIN116	0.93	0.07	0.40	TRAIN046	0.74	-0.21	0.07	TRAIN100	0.57	3.43	0.35
TRAIN023	0.91	-0.04	0.52	TRAIN049	0.74	-0.21	0.07	VALID015	0.57	17.83	0.18
VALID009	0.91	-0.05	0.49	TRAIN111	0.73	0.30	-0.24	TRAIN032	0.56	14.31	0.24
TRAIN013	0.91	-0.07	0.48	VALID018	0.70	-0.21	0.07	TRAIN111	0.54	18.24	0.11
TRAIN032	0.90	-0.06	0.50	TRAIN123	0.70	0.30	-0.25	VALID041	0.54	21.17	0.03
VALID000	0.90	-0.09	0.46	TRAIN018	0.68	-0.23	0.00	TRAIN114	0.53	-2.14	0.42
TRAIN017	0.89	-0.09	0.49	TRAIN004	0.65	-0.23	0.00	TRAIN102	0.53	3.12	0.33
TRAIN012	0.89	-0.08	0.48	VALID014	0.64	-0.23	-0.02	VALID007	0.53	18.29	0.16
TRAIN122	0.89	-0.08	0.48	VALID014	0.64	-0.23	-0.02	TRAIN122	0.51	1.03	0.33





Figure S7. Relative variation (%) of the 5 anthropogenic PPE (1 – 5) for the full period. Each pentagon represents the 5-D parameter space and the positions of the dots connected with lines show the position of each parameter within its range for that specific ensemble member. The filled area within the dots represents the explored parameter space in each ensemble member. Anticlockwise from top there are the five anthropogenic parameters: VBS\_AGERATE (P1), SVOC\_VOLDIST (P2), SVOC\_OXRATE (P3), IVOC\_SC (P4) and SVOC\_SC (P5). The values of the five parameters have been normalised dividing by their respective maximum values, hence their values in this plot range from 0 – 1. The colour in the lines and dots represents the FAC2 values from the O:C analysis and the fill colour represents the FAC2 values from the OA analysis. Red = 0 – 0.2, orange = 0.2 – 0.4, yellow = 0.4 – 0.6, green = 0.6 – 0.8, light blue = 0.8 -0.9 and blue = 0.9 -1.0

## S5. Emulator

### S5.1 Selection of periods for building and testing the emulator.

#### S5.1.1 Emulators for the four periods (high and low OA concentrations)

Four period time-slots were considered to build and test the emulator based on OA-AMS concentrations; one period with high OA concentrations (P1) and one period with low OA concentrations (P2). and two time-slots; 13:00 – 16:00 (A) and 20:00 – 23:00 hrs (B).

- 13:00-16:00 hrs (P1A) and 20:00-23:00 hrs (P1B), on 10 – 12 May, with high org mass loadings
- 13:00-16:00 hrs (P2A) and 20:00-23:00 hrs (P2B), on 20 – 23 May, with low org mass loadings

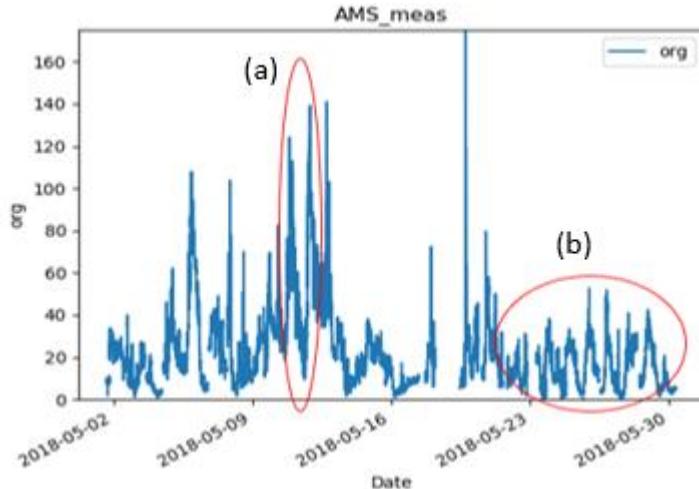


Figure S8. Time series of OA concentrations measured with the AMS highlighting the period with high (a) and low (b) Org concentrations.

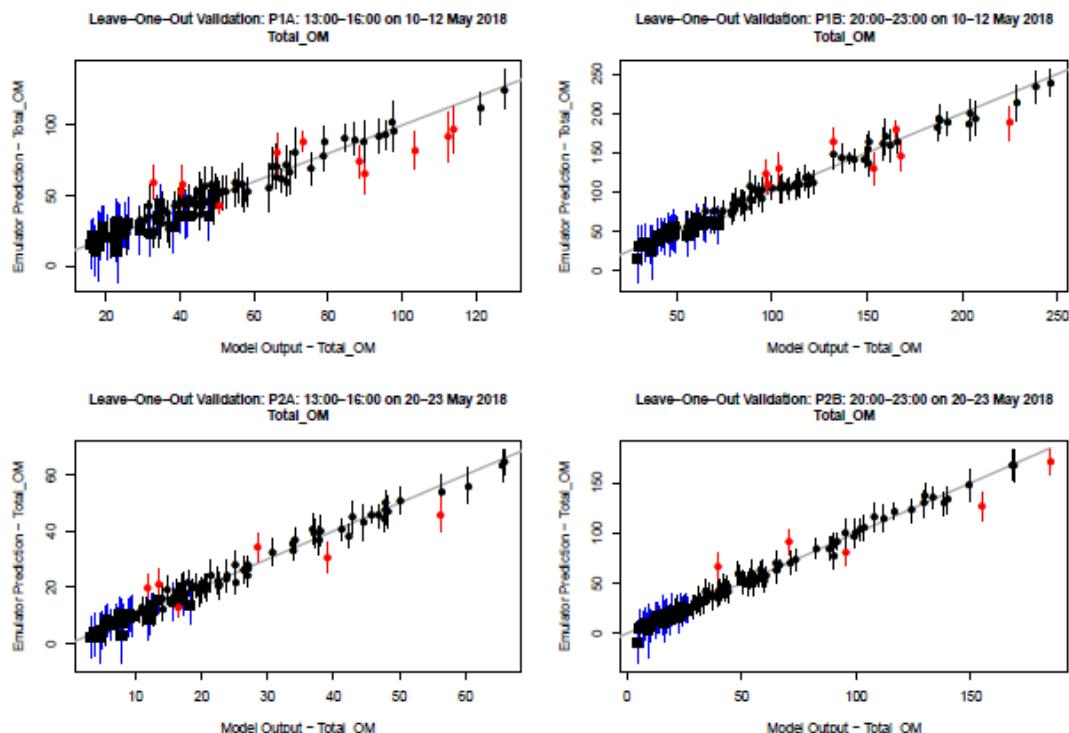


Figure S9. Validation of the four tested emulators for Org concentrations. Circles are the original 81 runs. Squares with error bars in blue are the new 30 runs with low settings of the anthropogenic SVOC scaling parameter (which has led to low aerosol mass). Runs where the actual model output lies outside the 95% prediction interval of the emulator are shown in red.

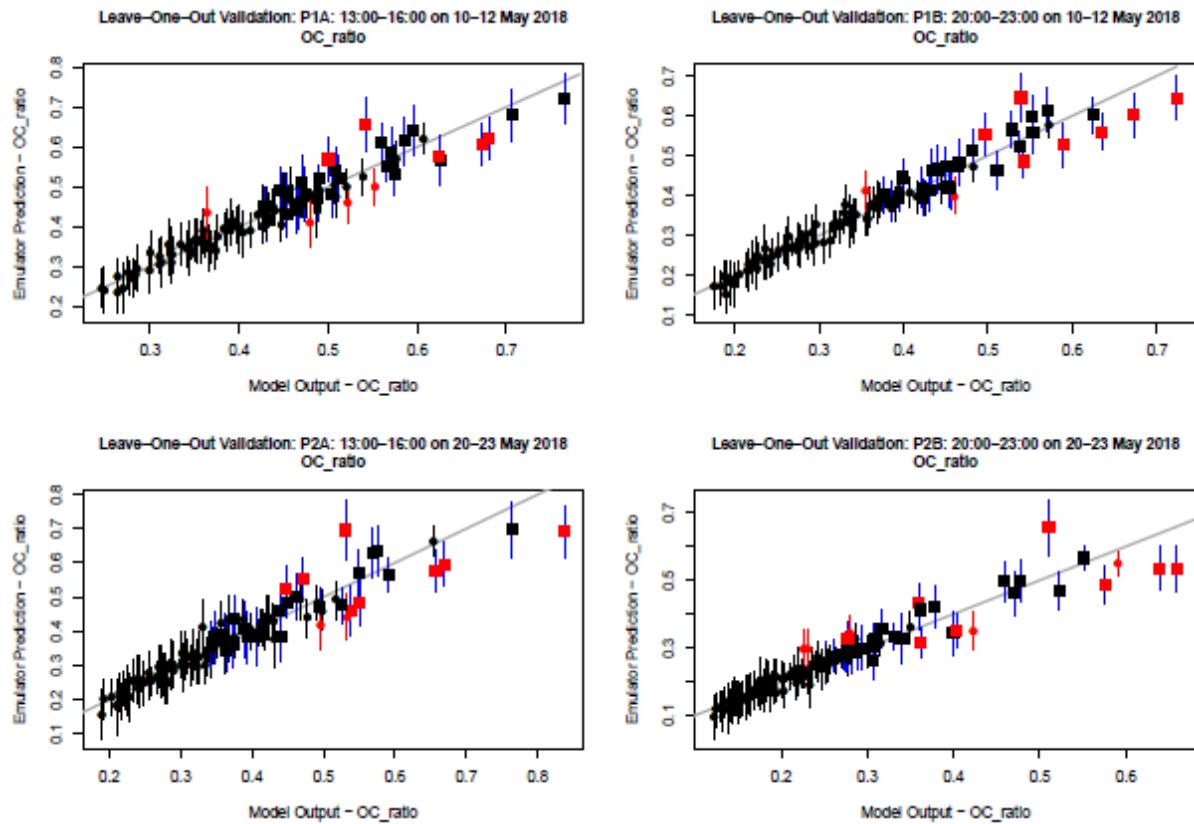


Figure S10 Validation of the four tested emulators for O:C ratios. Circles are the old 81 runs. Squares with error bars in blue are the new 30 runs with low aerosol mass. Red are the runs that are not within the 95% CI from prediction

### S5.1.2 Emulator for the 2-4 pm period

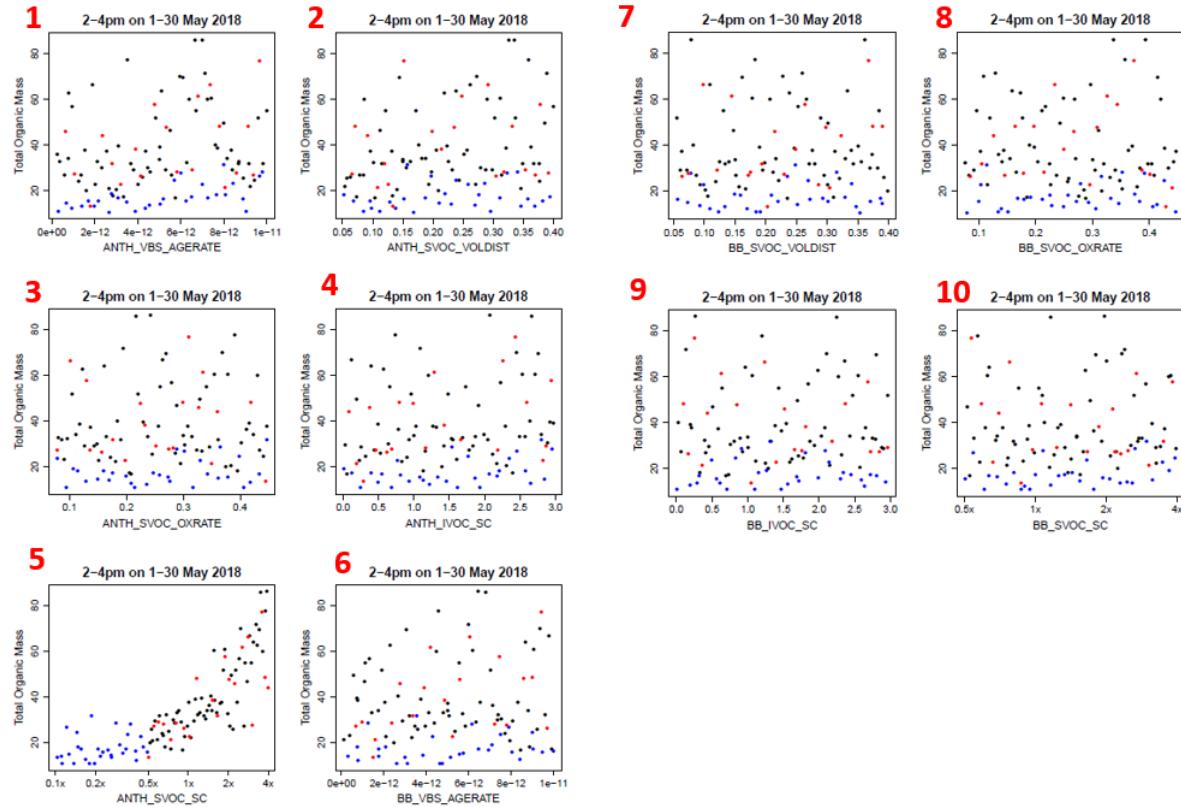


Figure S11 Spread of the total Organic mass for the 111 model runs vs the 10 parameters for the period 2-4 pm period. Red = 20 VALIDATE runs. Black = 61 TRAIN runs. Blue = 30 new TRAIN runs.

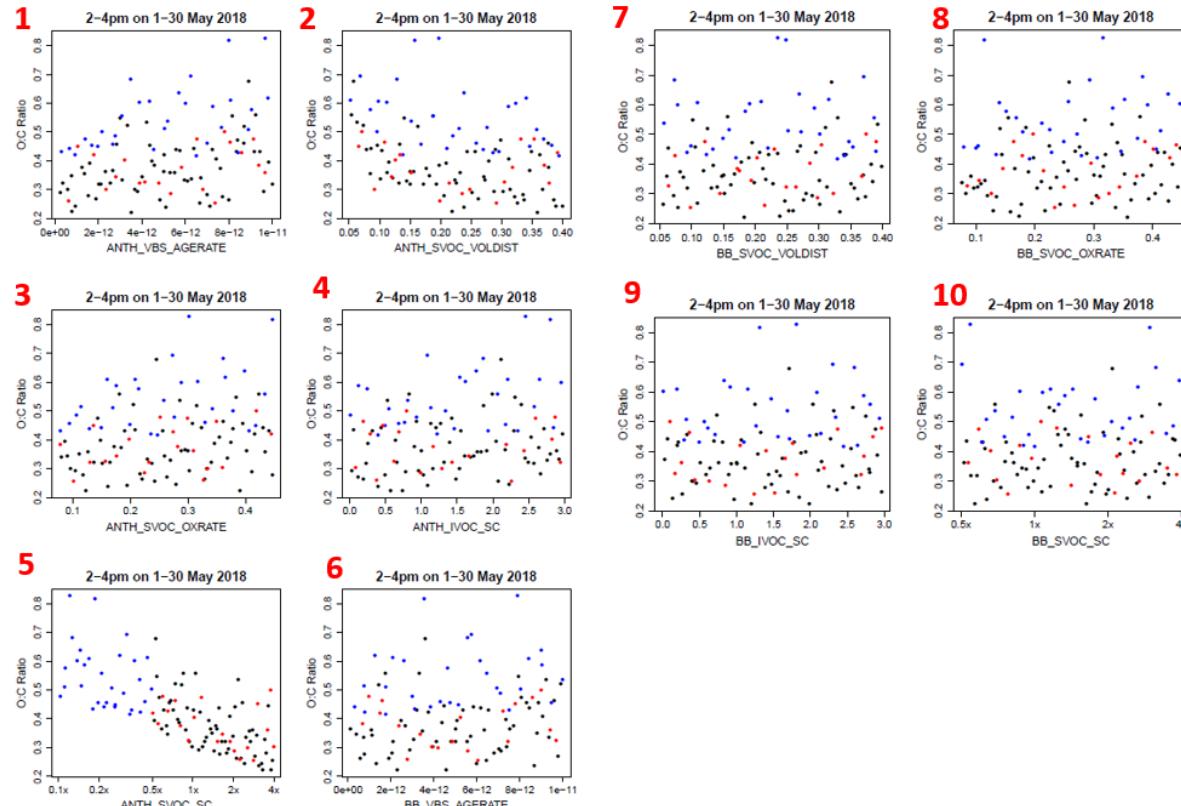


Figure S12. Spread of the total O:C ratio for the 111 model runs vs the 10 parameters for the 2-4 pm period. Red = 20 VALIDATE runs. Black = 61 TRAIN runs. Blue = 30 new TRAIN runs.

## S5.2 Total\_OM and O:C 2-d histogram analysis with and without constraint

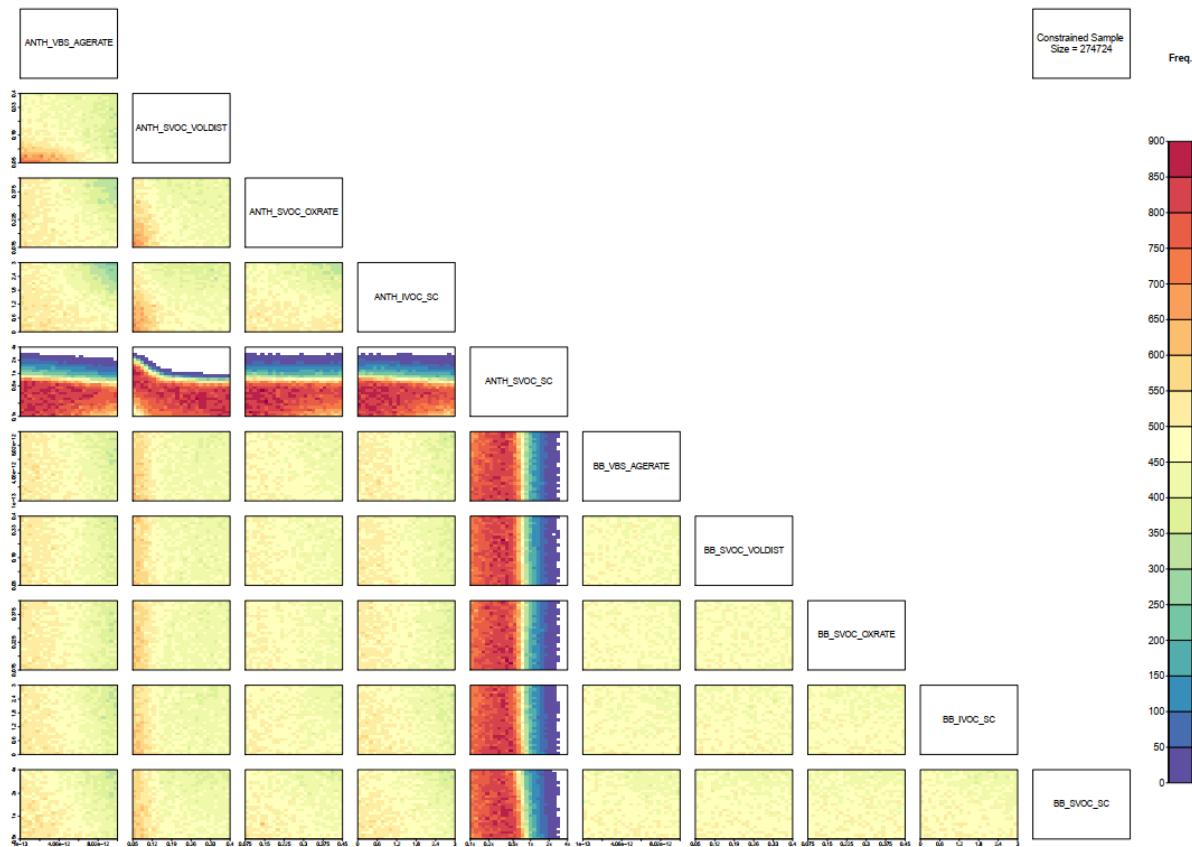


Figure S13. 2-d histogram for joint constraint effect (Total\_OM and OC\_ratio) for the full period accounting for emulator uncertainty. Retain 274724 variants from 0.5 million (~54.94%).

Carslaw, D. C., and Ropkins, K.: openair - An R package for air quality data analysis, Environ Modell Softw, 27-28, 52-61, DOI 10.1016/j.envsoft.2011.09.008, 2012.

Emmons, L., Pfister, G., Hodzic, A.: Information for mapping DESM2 (CAM-chem or WACCM) gas and aerosol species to WRF-Chem species for various mechanisms, report, accessible at:

[https://www2.acm.ucar.edu/sites/default/files/documents/CESM-WRFchem\\_aerosols\\_20190822.pdf](https://www2.acm.ucar.edu/sites/default/files/documents/CESM-WRFchem_aerosols_20190822.pdf), 2019.

Tsimpidi, A. P., Karydis, V. A., Zavala, M., Lei, W., Molina, L., Ulbrich, I. M., Jimenez, J. L., and Pandis, S. N.: Evaluation of the volatility basis-set approach for the simulation of organic aerosol formation in the Mexico City metropolitan area, Atmos. Chem. Phys., 10, 525-546, 10.5194/acp-10-525-2010, 2010.

Olivier, J., Peters, J., Granier, C., Pétron, G., Muller, J. F., and Wallens, S.: POET inventory - Metadata, GEIA-ACCENT emission data portal, France, 2003.

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, International Journal of Climatology, 32, 2088-2094, <https://doi.org/10.1002/joc.2419>, 2012.