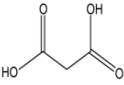
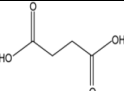
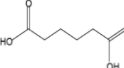


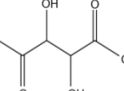
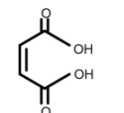
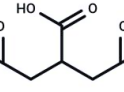
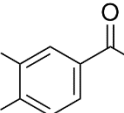
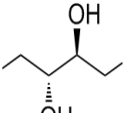
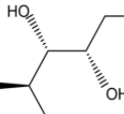



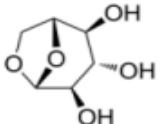
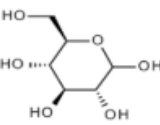
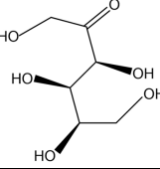
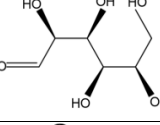
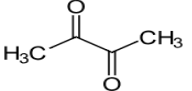
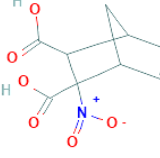
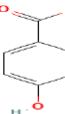
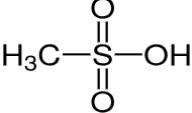
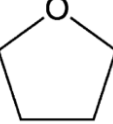
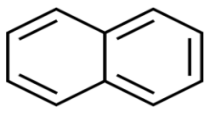
# From Single Compounds to Ambient Aerosols: A Machine-Learning-Based Estimation of Organic Hygroscopicity

S. Deshmukh et al.

## Supplementary findings.

5 Table S1. List of substances and their relevant properties investigated in this study.

Molecular Structure	Compounds	Chemical formula	Molar weight (g mol <sup>-1</sup> )	Density (g cm <sup>-3</sup> )	Solubility (g mL <sup>-1</sup> )	O/C	K <sub>measured</sub> (90% RH) *This work
	Malonic acid <sup>1</sup>	C <sub>3</sub> H <sub>4</sub> O <sub>4</sub>	104.06 <sup>a</sup>	1.62 <sup>a</sup>	0.763 <sup>b</sup>	1.33	0.28* 0.26 <sup>d</sup> 0.39 <sup>g</sup> 0.52 <sup>g</sup>
	Succinic acid <sup>1</sup>	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	118.09 <sup>a</sup>	1.19 <sup>a</sup>	0.083 <sup>b</sup>	1	0.01* 0.002 <sup>g</sup>
	Adipic acid <sup>1</sup>	C <sub>6</sub> H <sub>10</sub> O <sub>4</sub>	146.14 <sup>a</sup>	1.36 <sup>a</sup>	0.03 <sup>b</sup>	0.67	0.001* -0.003 <sup>g</sup>
	Pimelic acid <sup>1</sup>	C <sub>7</sub> H <sub>12</sub> O <sub>4</sub>	160.17 <sup>a</sup>	1.33 <sup>a</sup>	0.05 <sup>b</sup>	0.57	0.01* 0.03 <sup>g</sup>
	Suberic acid <sup>1</sup>	C <sub>8</sub> H <sub>14</sub> O <sub>4</sub>	174.19 <sup>a</sup>	1.3 <sup>a</sup>	0.0006 <sup>a</sup>	0.5	0.01* -0.03 <sup>g</sup>
	Tartaric acid <sup>1</sup>	C <sub>4</sub> H <sub>6</sub> O <sub>6</sub>	150.09 <sup>a</sup>	1.79 <sup>a</sup>	1.43 <sup>d</sup>	1.5	0.18* 0.22 <sup>g</sup>
	Maleic acid <sup>1</sup>	C <sub>4</sub> H <sub>4</sub> O <sub>4</sub>	116.07 <sup>a</sup>	1.59 <sup>a</sup>	0.79 <sup>a</sup>	1	0.19* 0.30 <sup>g</sup> 0.23 <sup>g</sup>
	Tricarballic acid <sup>1</sup>	C <sub>6</sub> H <sub>8</sub> O <sub>6</sub>	176.12 <sup>a</sup>	1.37 <sup>a</sup>	0.05 <sup>a</sup>	1.33	0.12*
	Vanillic acid <sup>1</sup>	C <sub>8</sub> H <sub>8</sub> O <sub>4</sub>	168.12 <sup>a</sup>	1.4 <sup>a</sup>	0.01 <sup>a</sup>	0.66	0.01*
	Meso-Erythritol <sup>1</sup>	C <sub>4</sub> H <sub>10</sub> O <sub>4</sub>	122.12 <sup>a</sup>	1.45 <sup>a</sup>	0.37 <sup>a</sup>	1.33	0.07*
	Xylitol <sup>1</sup>	C <sub>5</sub> H <sub>12</sub> O <sub>5</sub>	152.15 <sup>a</sup>	1.52 <sup>a</sup>	0.642 <sup>b</sup>	1	0.11*

	1,4-Butanediol <sup>1</sup>	C <sub>4</sub> H <sub>10</sub> O <sub>2</sub>	90.12 <sup>a</sup>	1.01 <sup>a</sup>	1.017 <sup>a</sup>	0.66	0.03*
	Levoglucosan <sub>1</sub>	C <sub>6</sub> H <sub>6</sub> O <sub>5</sub>	162.14 <sup>a</sup>	1.69 <sup>a</sup>	1.334 <sup>c</sup>	1.11	0.16* 0.18 <sup>i</sup> 0.165 <sup>h</sup>
	Glucose <sup>1</sup>	C <sub>6</sub> H <sub>12</sub> O <sub>6</sub>	180.15 <sup>a</sup>	1.54 <sup>a</sup>	0.909 <sup>f</sup>	1.33	0.13* 0.15 <sup>h</sup> 0.13 <sup>c</sup>
	D(-)fructose <sup>1</sup>	C <sub>6</sub> H <sub>12</sub> O <sub>6</sub>	180.16 <sup>a</sup>	1.59 <sup>a</sup>	3.75 <sup>a</sup>	1	0.09* 0.23 <sup>g</sup>
	D(+)mannose <sub>1</sub>	C <sub>6</sub> H <sub>12</sub> O <sub>6</sub>	180.16 <sup>a</sup>	1.54 <sup>a</sup>	2.48 <sup>a</sup>	1	0.08* 0.14 <sup>g</sup>
	2,3-Butanedione <sup>1</sup>	C <sub>4</sub> H <sub>6</sub> O <sub>2</sub>	86.09 <sup>a</sup>	0.001 <sup>a</sup>	0.981 <sup>a</sup>	0.5	0.01*
	Humic acid <sup>2</sup>	C <sub>9</sub> H <sub>9</sub> NO <sub>6</sub>	227.17 <sup>a</sup>	1.77 <sup>a</sup>	0.01 <sup>a</sup>	0.8	0.07*
	4-Hydrobenzoic acid <sup>1</sup>	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>	138.12 <sup>a</sup>	1.46 <sup>a</sup>	0.005 <sup>a</sup>	0.57	0.01*
	Methane Sulfonic acid <sup>1</sup>	CH <sub>4</sub> O <sub>3</sub> S	96.11 <sup>a</sup>	1.48 <sup>a</sup>	1.47 <sup>a</sup>	-	0.21*
	Tetrahydrofuran <sup>1</sup>	C <sub>4</sub> H <sub>8</sub> O	72.11 <sup>a</sup>	0.89 <sup>a</sup>	0.88 <sup>a</sup>	0.33	0.01*
	Naphthalene <sup>1</sup>	C <sub>10</sub> H <sub>8</sub>	128.17 <sup>a</sup>	1.16 <sup>a</sup>	0.00003 <sup>a</sup>	0	0.17*

Supplier purity, <sup>1</sup>Sigma Aldrich, ≥ 99 %, <sup>2</sup>Fluka, 98%. <sup>a</sup><https://www.chemicalbook.com/> (last access: 05 December 2025).

<sup>b</sup><https://pubchem.ncbi.nlm.nih.gov/> (last access: 05 December 2025). <sup>c</sup><https://www.chemspider.com/> (last access: 05

December 2025). <sup>d</sup>Peng et al. (2001). <sup>e</sup>Lei et al. (2023). <sup>f</sup>Chan et al. (2008). <sup>g</sup>Previous literature κ values used from reference

Table 2, S. Han et al. (2022). <sup>h</sup>Koehler et al. (2006). <sup>i</sup>Mochida et al. (2004)

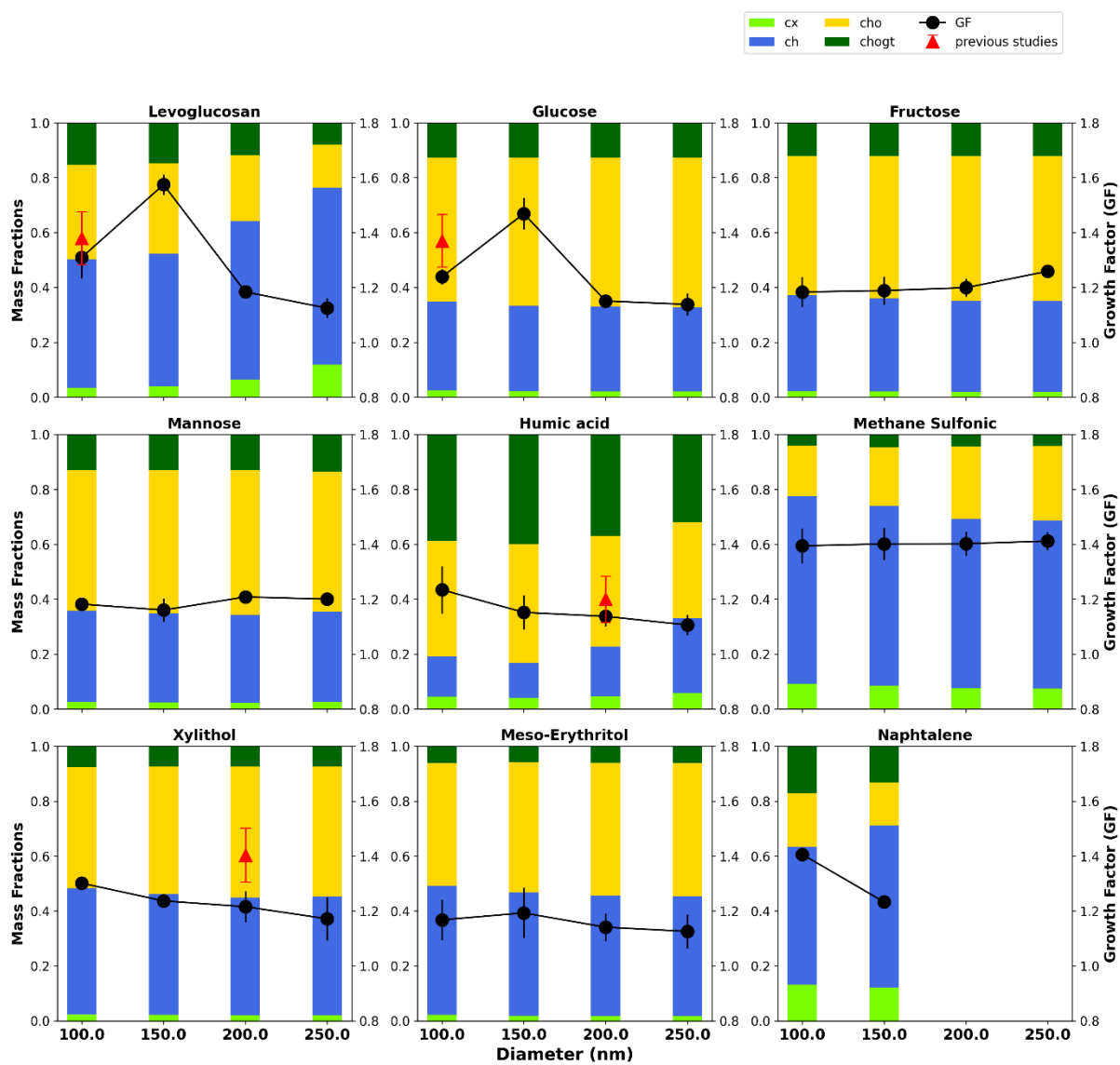


Figure S1: The family groups (Cx, CH, CHO, CHOgt1) mass fractions and hygroscopic growth values of straight-chain dicarboxylic acid particles for different diameters were measured in this study, and their comparison with previous studies

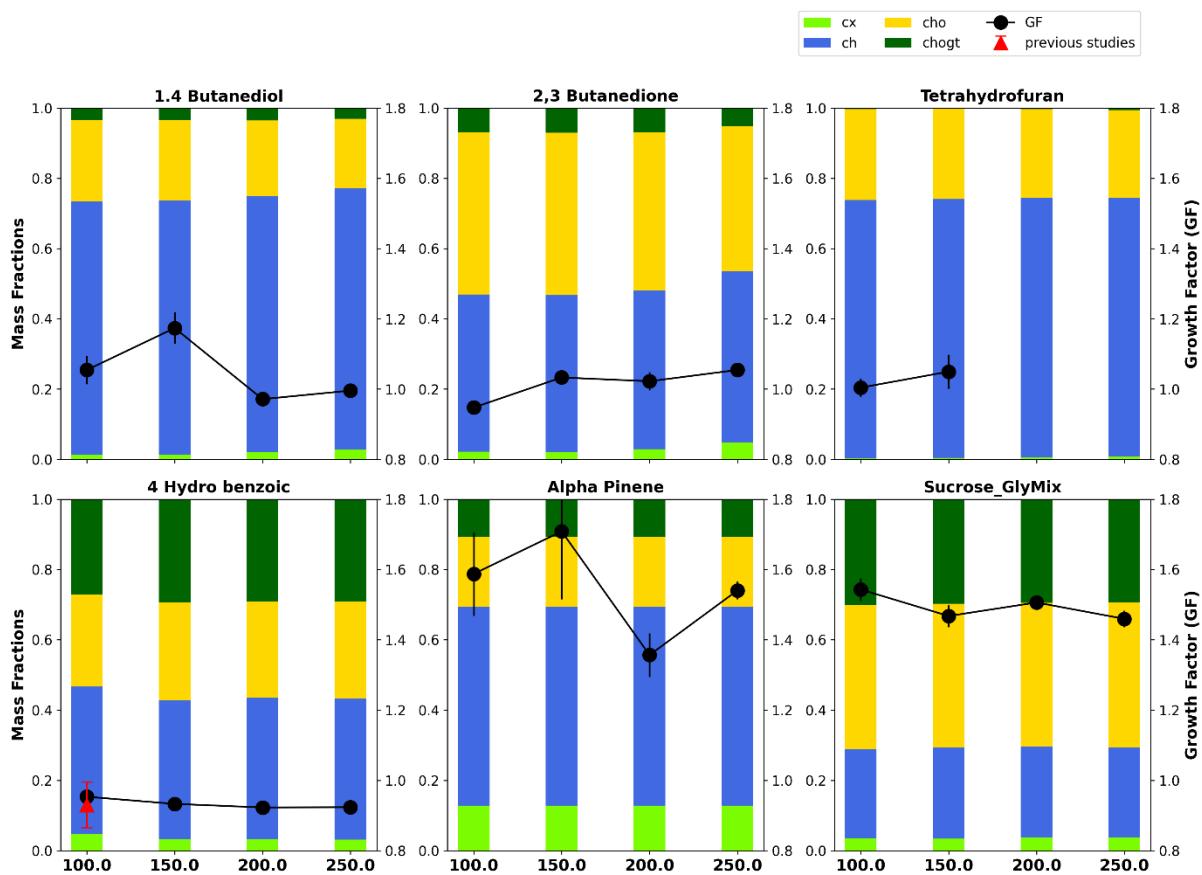
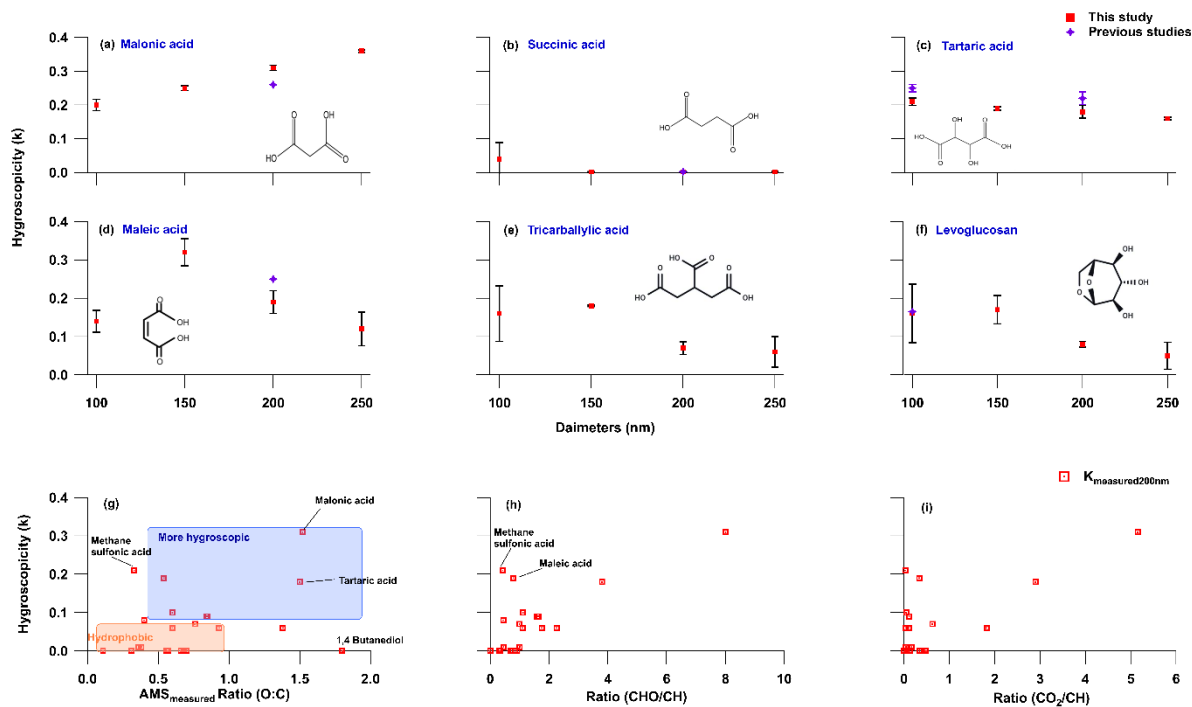
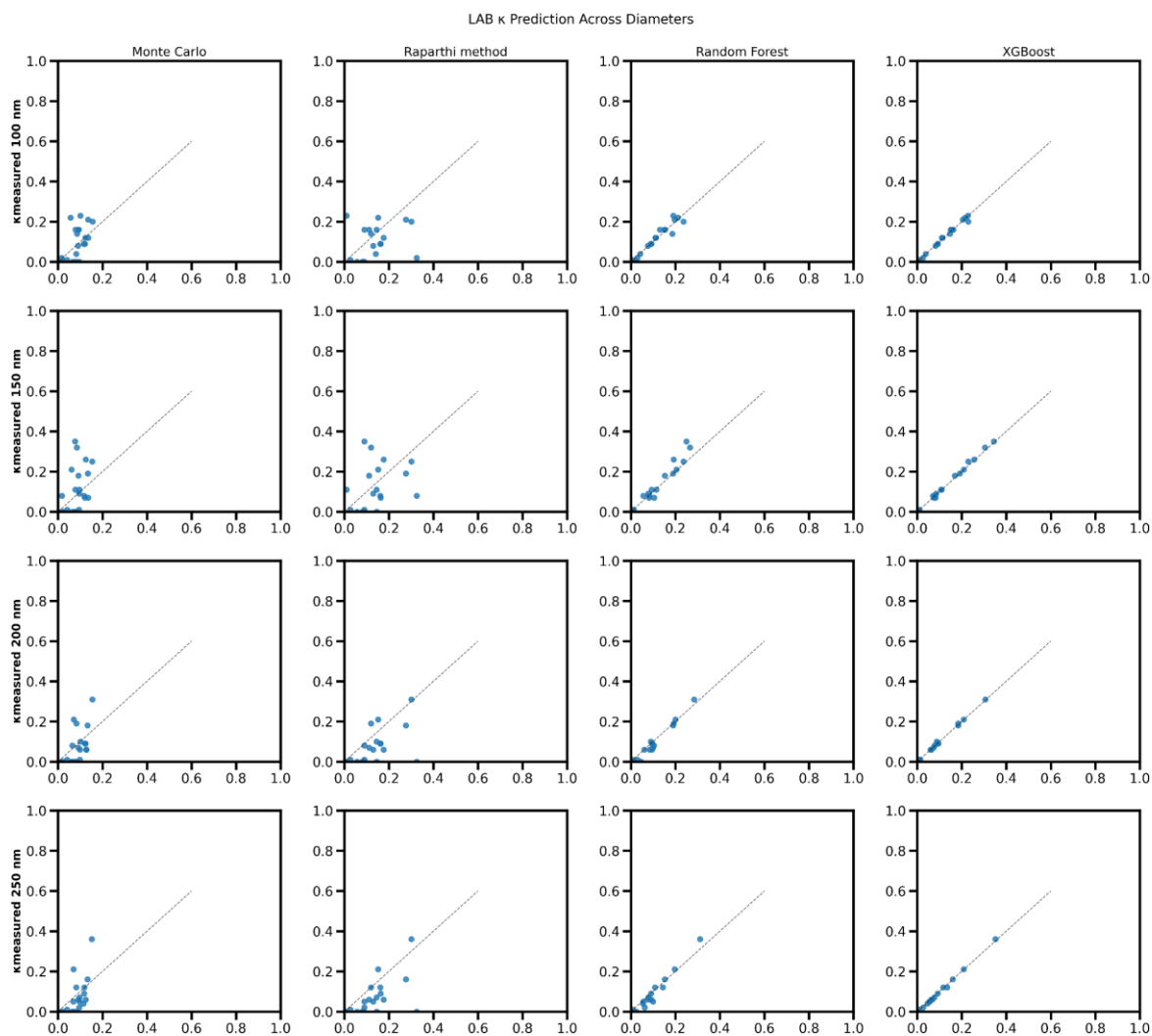


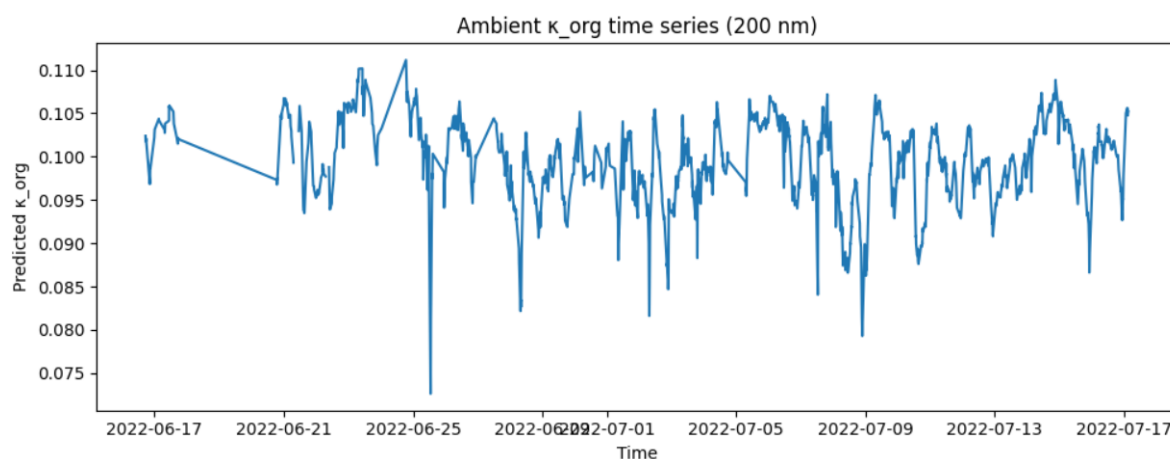
Figure S2: The family groups (Cx, CH, CHO, CHOgt1) mass fractions and hygroscopic growth values of straight-chain dicarboxylic acid particles for different diameters were measured in this study, and their comparison with previous studies.



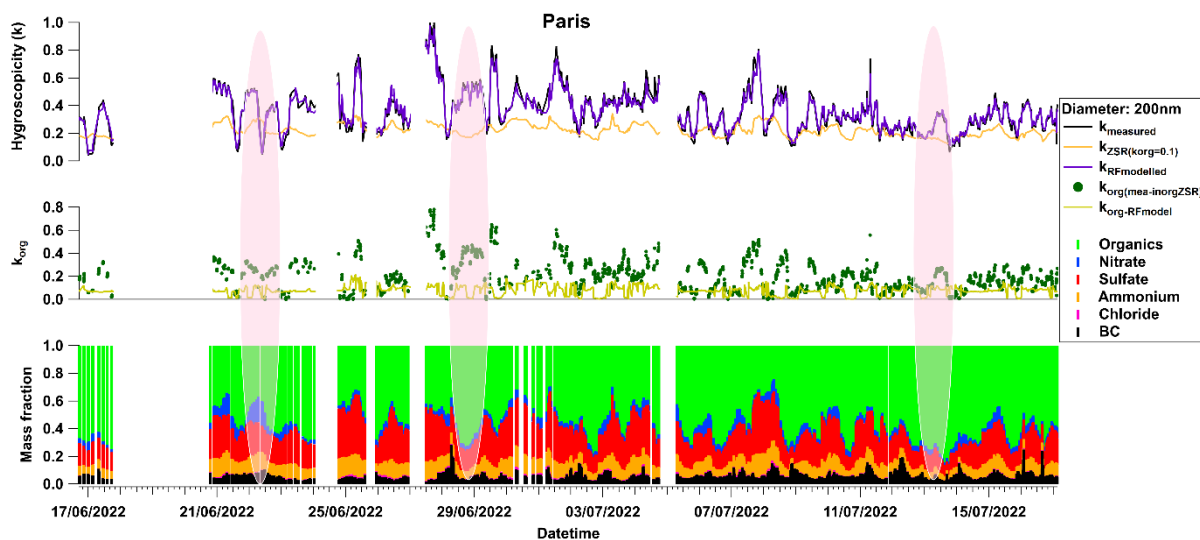
20 Figure S3: The hygroscopic growth ( $\kappa_{org}$ ) variation of organic compounds for different diameters, measured in this study, and compared with previous literature (a-f).  $\kappa_{org}$ 's relation to the measured O:C ratio from AMS and comparison for CHO and CH structures for 200nm with other diameters in Fig S7 and the supplementary supporting files.



25 **Figure S4: Estimation of  $\kappa_{org}$  using Monte Carlo for laboratory standard organic compounds using family group kappa classifications, Monte Carlo, and ML models at different diameters.**

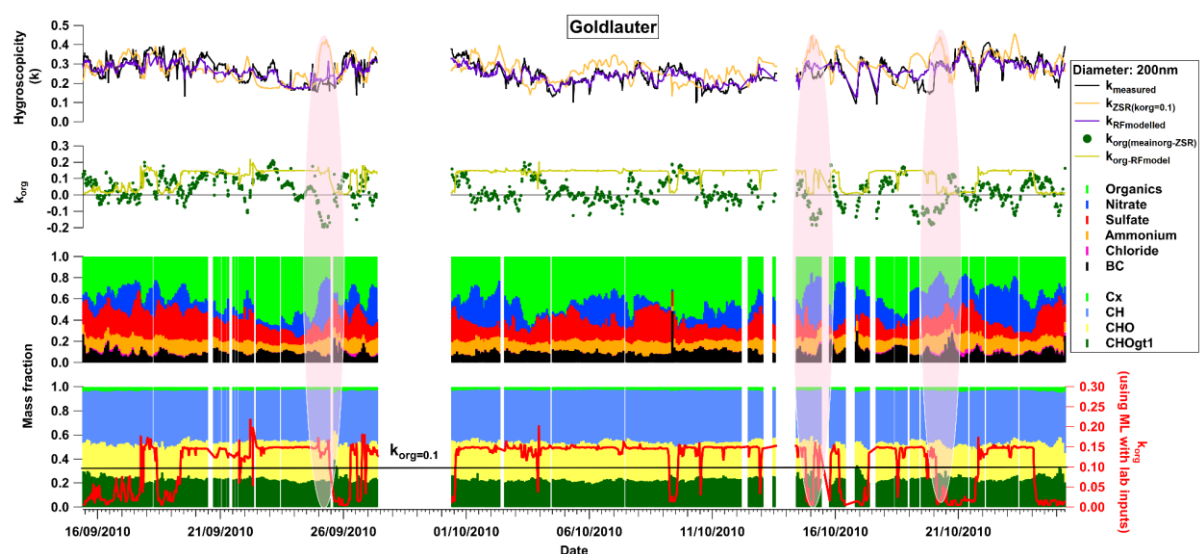


**Figure S5: Estimation of  $\kappa_{org}$  using the RF model with pre-test laboratory standard organic compound using family group kappa classifications. Testing on ambient measurements from Paris at 200nm.**



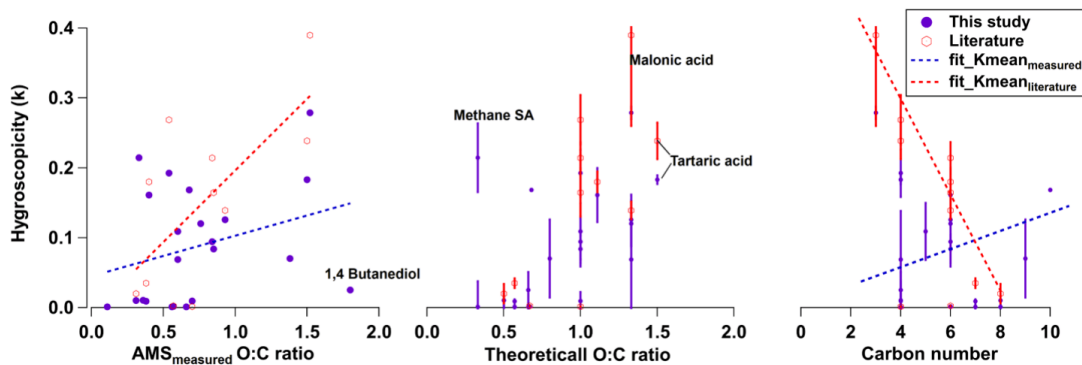
30

Figure S6: Comparison of  $\kappa$  from ZSR and ML output (RF) for 200nm with variation of  $\kappa_{\text{org}}$  in both methods at Paris measurement. Shaded regions indicate events with high organic-mass fractions. For other diameters, the outputs are in supplementary files.



35

Figure S7: Comparison of  $\kappa$  from ZSR and ML output (RF) for 200nm with variation of  $\kappa_{\text{org}}$  in both methods at Goldlauter measurement. Shaded regions indicate events with high organic-mass fractions. For other diameters, the outputs are in supplementary files.



Compound	Kmean thiswork	SD	Kmean_lit	SD	Ratio_OC	Therotical_OC	Carbon Number
Malonic	0.278813958	0.0089417	0.39	0.13	1.52	1.33	3
Succinic	0.009299852	0.0128527	0.0015	0.0007071	0.7	1	4
Adipic	0.001622189	0.0006153	0.0025	0.0007071	0.57	0.67	6
Tartaric	0.183006991	0.0061137	0.2385	0.026163	1.5	1.5	4
Pimelic	0.009254055	0.0026958	0.035	0.0070711	0.38	0.57	7
Suberic	0.01	0.0085466	0.02	0.0141421	0.31	0.5	8
Maleic	0.192593935	0.0343255	0.26866667	0.0355715	0.54	1	4
Vanillic	0.001	0.0051363			0.66	0.66	8
Tricarballic	0.120228919	0.032664			0.76	1.33	6
1,4 Butanediol	0.025031696	0.0259572			1.8	0.66	4
2,3 Butanedione	0.010545132	0.0113523			0.36	0.5	4
Tetrahydrofuran	0.00087935	0.0367771			0.11	0.33	4
4 Hydrobenzoic	0.001	0.0034148			0.56	0.57	7
Levogluconan	0.161019629	0.0388532	0.18	0.015	0.4	1.11	6
Glucose	0.125699584	0.0358455	0.139	0.0127279	0.93	1.33	6
Fructose	0.09430123	0.0358038	0.214	0.0226274	0.84	1	6
Mannose	0.08385672	0.020167	0.1645	0.0346482	0.85	1	6
Humic acid	0.070255187	0.0558627		0.005	1.38	0.8	9
Methane Sulfonic acid	0.214600225	0.0494163		0.001	0.33	0.33	
Xylithol	0.108872738	0.0409536		0.007	0.6	1	5
Meso-Erythritol	0.068735949	0.069378		0.014	0.6	1.33	4
Naphtalene	0.168301687	0.0013129		0.01	0.68	0.68	10

Figure S8: Comparison of  $\kappa_{\text{measured}}$  from this study to the Literature with the O:C ratio and Carbon number. Here,  $\kappa_{\text{mean}}$  is the average of all diameters (100, 150, 200, and 250nm), with the standard deviation for this work, referred to as bulk  $\kappa_{\text{measured}}$ . This is similar to  $\kappa_{\text{mean-lit}}$ , which is the average of all previous studies' measurements for available compounds. For  $\kappa_{\text{mean-lit}}$  data points, refer to Table S1.

40

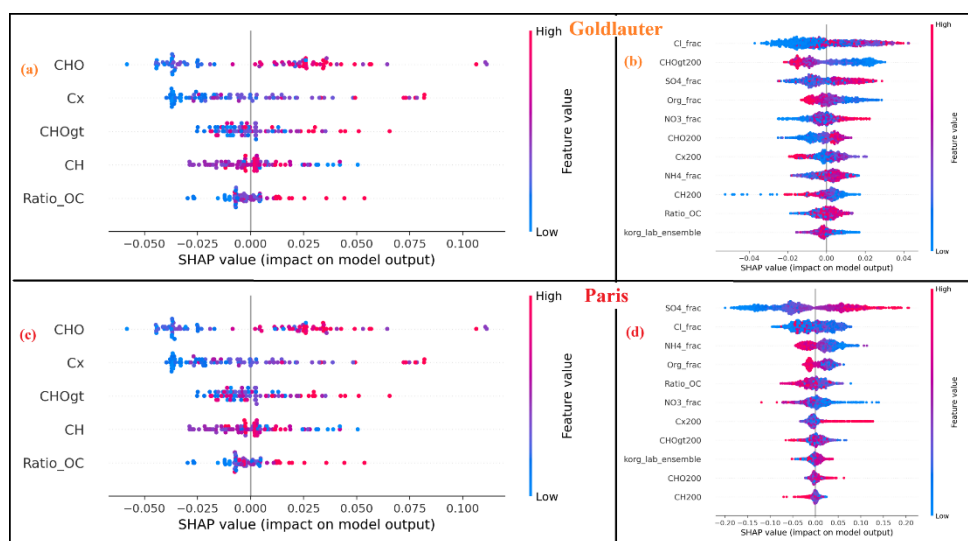


Figure S9: Model interpretability was analysed using permutation importance and Shapley additive explanations value diagnostics, confirming that oxygenation level (O:C), CHOgt, CH family groups, and inorganic fractions are the dominant predictors for  $\kappa_{\text{org}}$  and  $\kappa_{\text{measured}}$ , respectively. (a & c) The laboratory SHAP interpretation, similarly using ambient data (b & d), is shown for the Goldlauter and Paris sides in the top and bottom panels, respectively.

45