

Interactive comment on ‘From Single Compounds to Ambient Aerosols: A Machine-Learning-Based Estimation of Organic Hygroscopicity’ by S. Deshmukh et al.,

We thank the reviewers for their constructive and valuable comments, which has helped to improve the quality of our manuscript. We have considered all the reviewers' comments and made the necessary changes to the paper. The details of our changes are highlighted in the text and colour-coded. The point-by-point answers to all the reviewers are provided below.

Color code: Black: comment from reviewer; Blue: Reply/Response and sentence from the manuscript; Brown: Change/addition in the revised manuscript.

Response to Referee comments #RC1

1. [General Comment] The manuscript "From Single Compounds to Ambient Aerosols: A Machine-Learning-Based Estimation of Organic Hygroscopicity" by Deshmukh et al. provides estimates of the hygroscopicity parameter κ for single-component organic aerosols, and uses these hygroscopicity parameters to build a model to predict the hygroscopicity of ambient aerosol using functional groups that can be measured by aerosol mass spectrometry. Overall, this seems like a promising approach to predict the hygroscopicity of ambient aerosol. However, the manuscript text and figures need substantial revision to improve its readability.

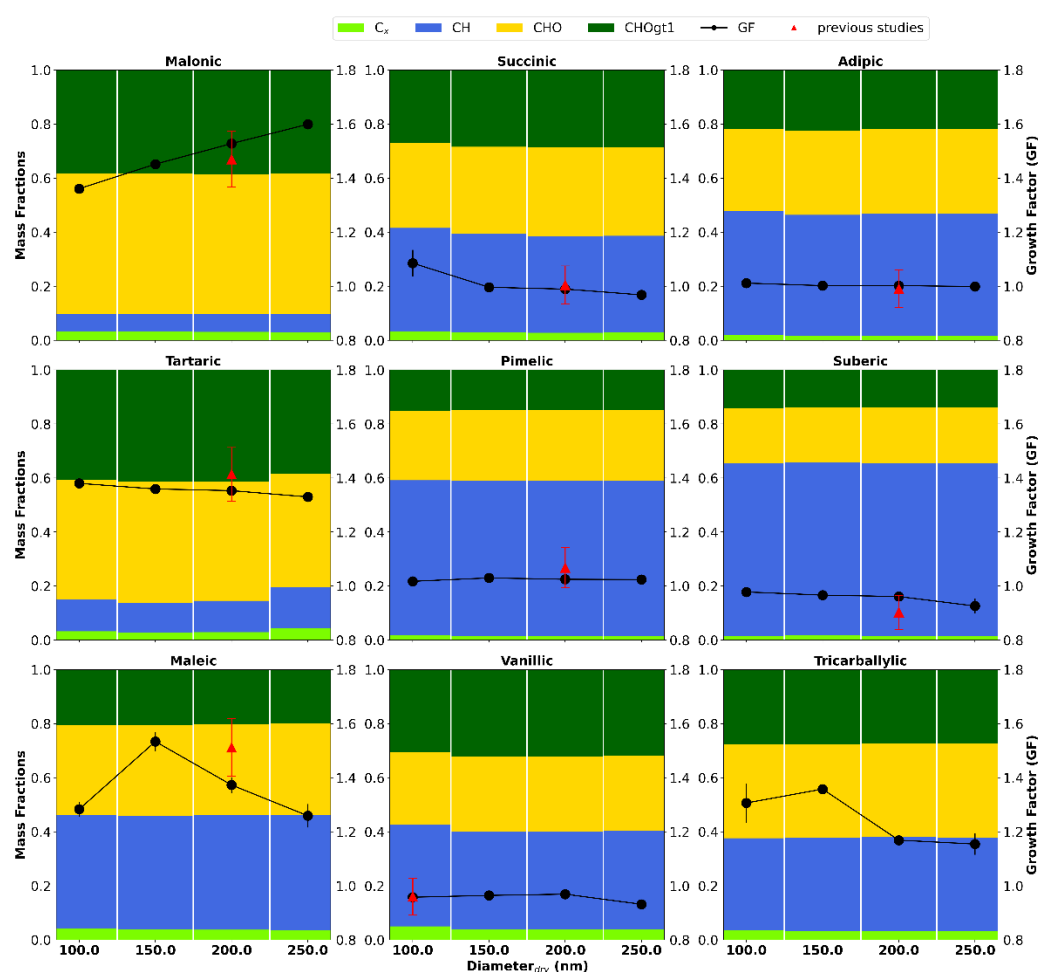
Response: We thank the reviewer for the constructive comment and positive summary of our work. Below we list a point-by-point response to all questions and suggestions that improved the paper.

2. [General Comment] Presentation of Figures:

- Fig 2 - I believe the x-axis is dry diameter? Can the authors please clarify this.
- Fig 2 - It is not clear why i) the error bars for the literature data are so large compared to this work, ii) where the literature data came from
- Fig 2 - the caption says straight-chain dicarboxylic acid particles but there are molecules that do not seem to fit this description in the panes.
- Fig 2- CHO_{gt} has never been defined.

Response: Thank you. We have revised the figure and caption as follows. First, the x-axis now explicitly states “Dry particle diameter, D_{dry} (nm). Second, the literature data are reported in Table S1, and the meaning of the error bars is explicitly described. The comparatively large variability in the literature values reflects [the spread among values reported in different studies, in short, the range of reported values], whereas the error bars for our measurements represent [state exactly: replicate variability, standard deviation].

These quantities are therefore not necessarily directly equivalent. Third, we replaced “straight-chain dicarboxylic acids” with “selected carboxylic acids,” because Fig. 2 also includes substituted dicarboxylic, aromatic carboxylic, and tricarboxylic acids. Finally, CHOgt1 is the family label used in our AMS data-processing workflow, where “gt1” denotes “greater than one.” It represents the summed oxygenated organic-ion family $C_xH_yO_z^+$, $z > 1$. We retained this label in the figures to remain consistent with the AMS family notation used in the analysis, but we now define it explicitly at its first occurrence in Section 2.3 and in the captions of Figs. 1 and 2. We have also standardized the notation to “CHOgt1” throughout the manuscript and Supplement. We clarify that these categories represent AMS ion-fragment families rather than directly identified molecular functional groups. This notation is now defined at its first occurrence in Section 2.3 and in the relevant figure captions.



- [Lines 375-382] “...Figure 2: Size-segregated hygroscopicity of selected dicarboxylic acid particles measured at 90 % RH, together with AMS-relevant compositional proxies expressed as family groups (Cx, CH, CHO, CHOgt1) for different particle diameters. Whereas CHOgt1 denotes the summed AMS organic-ion family ($C_xH_yO_z^+$, $z > 1$, i.e., fragments containing more than one oxygen atom). Results are compared with available literature data to illustrate the influence of functionalization and particle size on organic hygroscopicity. The literature or

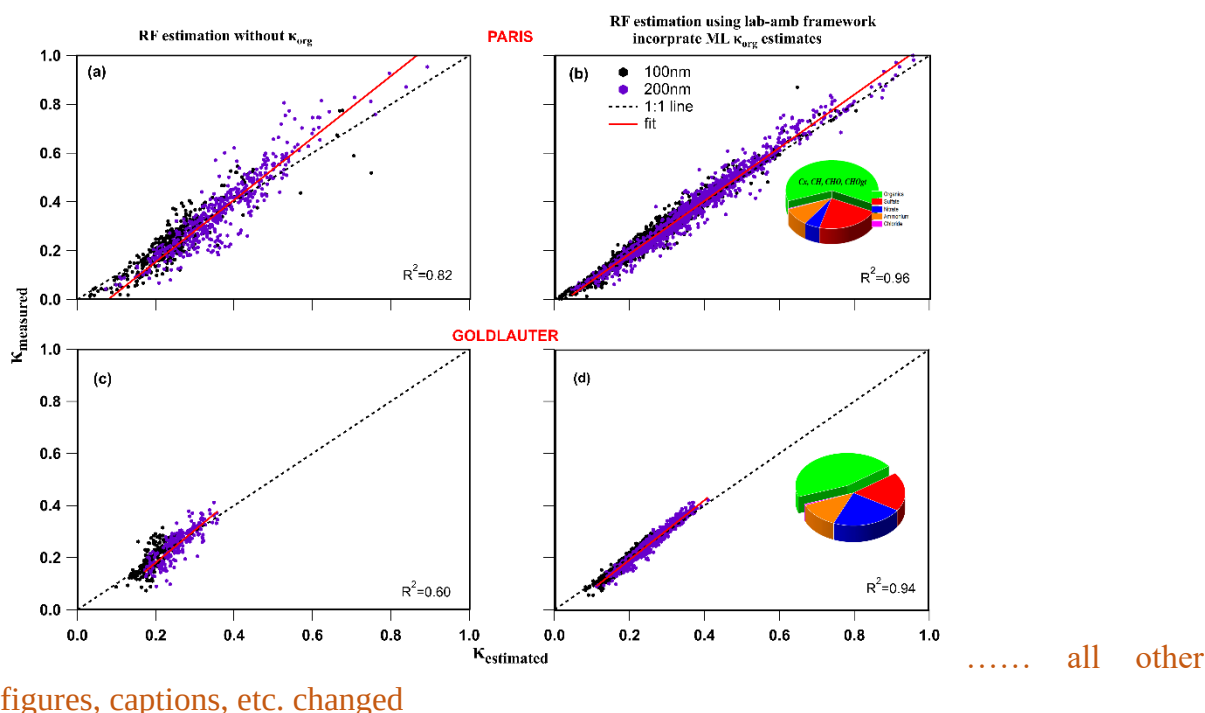
previous study error bars reproduce the spread or range among values reported in different studies, as reported in their original publications (refer to Table S1) ...”

- [Lines 220-225] “...represents the sum of the organic fragments having different combinations of C, H, and O that define the family. Organic ions were grouped according to their elemental composition into the AMS ion families C_x (C_x⁺), CH (C_xH_y), CHO (C_xH_yO⁺, containing one oxygen atom), and CHOgt1 (C_xH_yO_z⁺, z > 1, containing more than one oxygen atom). The abbreviation “gt1” denotes “greater than one” and follows the family notation used in the AMS data-processing workflow. These categories represent ion-fragment families rather than directly identified molecular functional groups...”

3. [General Comment] Presentation of Figures:

Fig 3/4 /5/6 use k rather than kappa for the hygroscopicity parameter. These also do not look consistent between them, sometimes a capital and sometimes not?

Response: We thank the reviewer for identifying this inconsistency. We have revised Figs. 3–6 and their captions to use the Greek symbol κ consistently for the hygroscopicity parameter. It was mistakenly replaced during plotting; it has been changed throughout the manuscript. [Please check the revised manuscript.](#)



figures, captions, etc. changed

4. [General Comment] Presentation of Figures:

Fig S1 and 2S both say straight-chained dicarboxylic acid particles in the caption, while none of these molecules fit into that category.

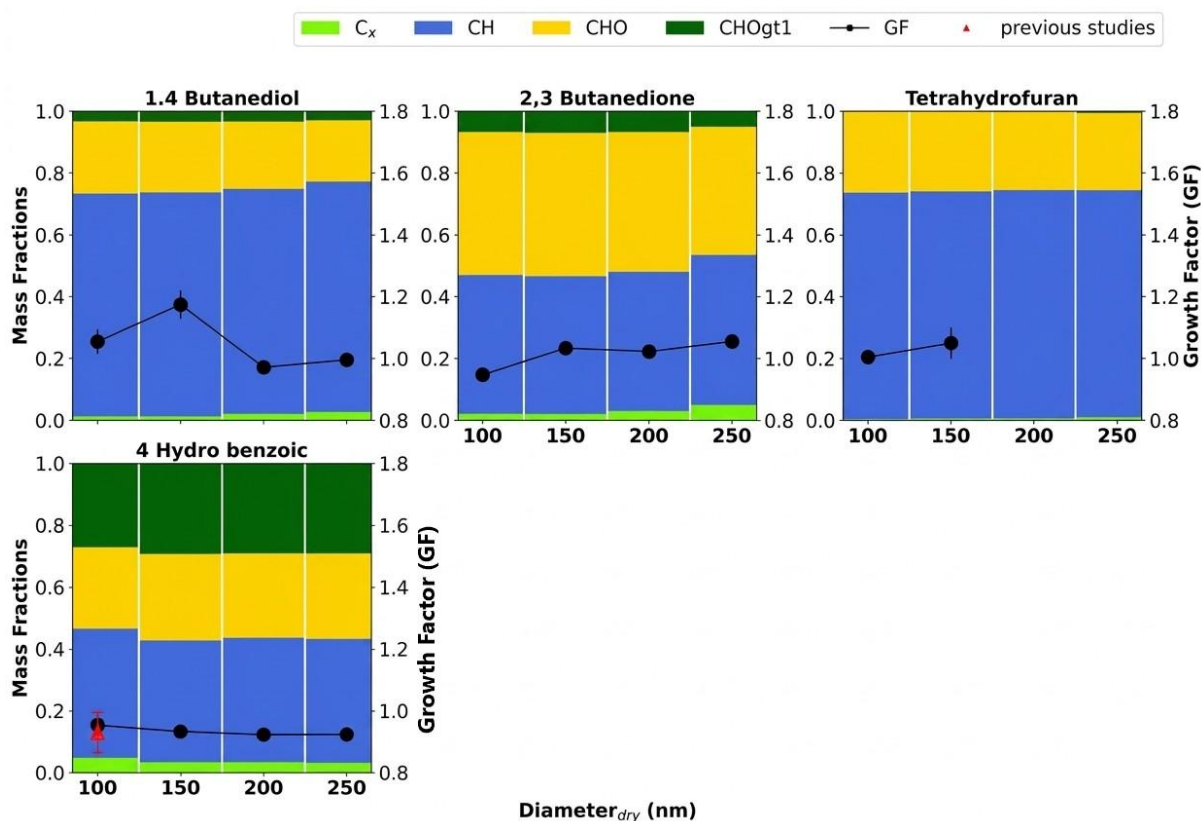
Response: We thank the reviewer for identifying this inconsistency. We have revised all the captions; they have been changed throughout the manuscript.

[Lines 858-861] “...Figure S1: The family groups (C_x, CH, CHO, CHOgt1) mass fractions and hygroscopic growth values of sugars and selected organic compounds for different diameters were measured in this study, and their comparison with previous studies...”

5. [General Comment] Presentation of Figures:

Between Fig 2 S1 and S2 there are more hygroscopicity plots than molecules in table S1. what is sucrose_gly mix in Fig S2? I do not think it is mentioned anywhere in the manuscript.

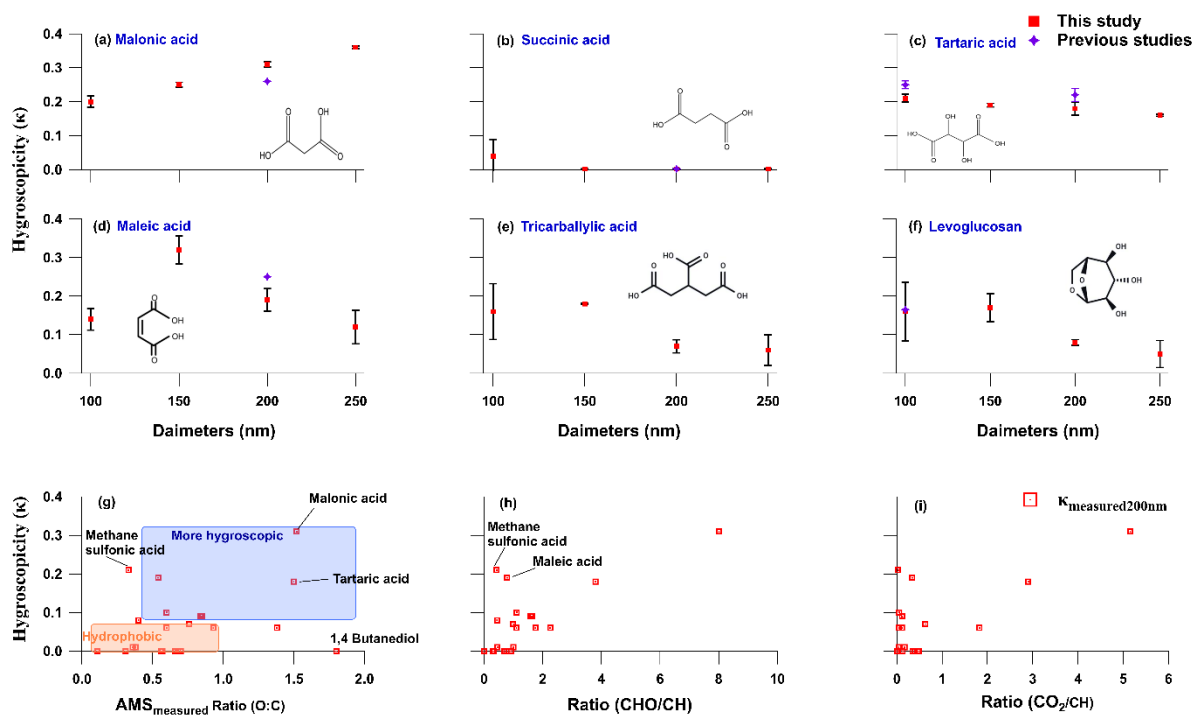
Response: We thank the reviewer for identifying this inconsistency. The sucrose–glycerol mixture was measured during the laboratory campaign as an exploratory additional sample but was not included in the set of 22 compounds (Table S1) or in any statistical, Monte Carlo, or machine-learning analyses. Because it is outside the scope of the dataset evaluated in this study and was not described elsewhere in the manuscript, we have removed it from Fig. S2. We have also cross-checked Figs. 2, S1, and S2 against Table S1 and revised the figures and captions to ensure that only compounds included in the reported dataset are presented.



6. [General Comment] Presentation of Figures:

Fig s3 - the axis labels and text in the figure is all far too small to see. The x-axis label of diameter on the a-f sub plots is not clear as the label is only in the middle panel while the bottom row of the figure has every panel labeled.

Response: We thank the reviewer for the suggestion. We revised the manuscript and figures.

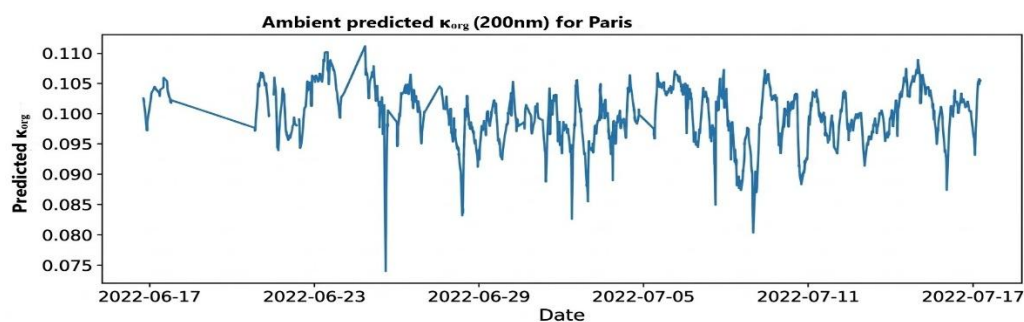


[General Comment] Presentation of Figures:

- Fig S4 - there are no x-axis labels included.
- Fig S5 - some of the x-axis labels are overlapping making them impossible to read.
- Fig S6/S7 - please indicate the meaning of the pink regions in the figure captions.
- Fig S9 - the text in panels b and d is too small to read.

Response: Thank you, we modified the figure in the revised manuscript. We also rechecked all figures in the main and supplementary (labels, axis ticks, etc.). Please check the revised manuscript for the changed figures.

- Figure S4: axis labels added and figure revised
- Figure S5: Axis changed and figure modified.



[Lines 795-804] “...Figure S6: Comparison of κ from ZSR and ML output (RF) for 200nm with variation of κ_{org} in both methods at Paris measurement. Shaded regions indicate events with high organic-mass fractions. The pink-shaded region shows a large discrepancy between κ_{measured} (black line) and κ_{ZSR} (orange line), due to variations in κ_{org} (green dots). Using $\kappa_{\text{org-RFmodel}}$ estimates improves κ prediction (purple line), which resembles κ_{measured} . For other diameters, the outputs are in supplementary files...”

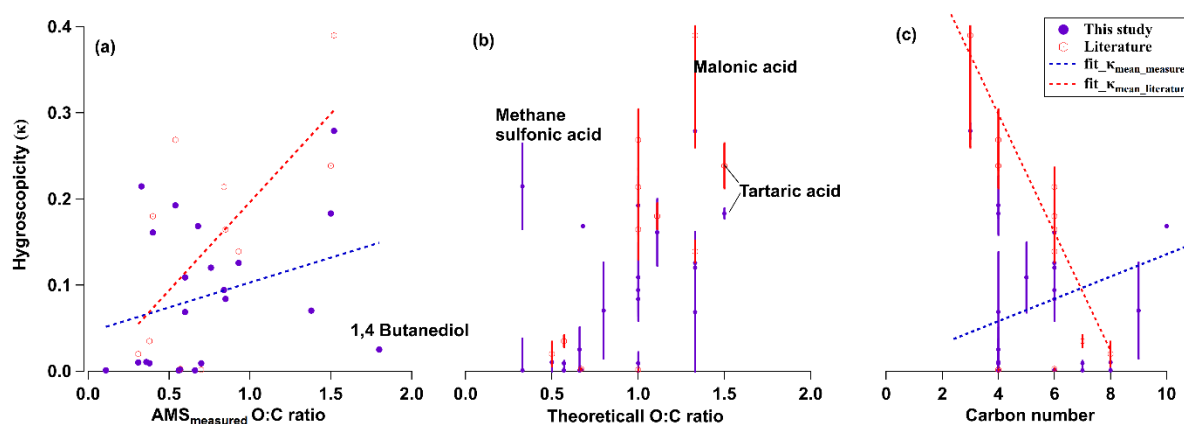
- Figure S9: Figure modified.

[General Comment] Presentation of Figures:

Fig S8 - only one of these subplots is referred to in the text. Is the table part of the figure? Please only report the significant values of kappa and its standard deviation. The standard deviation should never have greater precision than the reported value.

Response: We thank the reviewer for pointing out these important inconsistencies in presentation and formatting. We have fully overhauled the figure and table properties to comply with these recommendations:

1. Text References: We have revised the main manuscript text to explicitly refer to and discuss all three subpanels (hygroscopicity vs AMS-measured O:C ratio, theoretical O:C ratio, and carbon number) to ensure the entire multi-panel graphic is justified and properly integrated. [Refer to section 3.2.](#)
2. Table Separation: The embedded data overview has been extracted from the image file completely. It is now presented within the Supplementary Material file.
3. Significant Figures and Precision Matching: All values and their corresponding standard deviations have been rounded appropriately. Standard deviations are reported to two significant figures, and the corresponding mean value's precision has been truncated to match the exact same decimal place.



[Lines 805-806] “...Figure S8: Comparison of κ_{measured} from this study with AMS-measured O:C (a) and the literature. The O:C ratio and Carbon number (b-c). Here, κ_{mean} is the average of all diameters (100, 150, 200, and 250nm), with the standard deviation for this work, referred to as bulk κ_{measured} . This is similar to $\kappa_{\text{mean-lit}}$, which is the average of measurements from all previous studies for available compounds. For $\kappa_{\text{mean-lit}}$ data points, the error bars represent the spread or range of values reported in different studies; refer to Table S1...”

[General Comment] Clarity of the text:

line 16 - what do the authors mean by 'similar to carbon chains, increased functionalization generally enhances hygroscopicity'.

Response: We thank the reviewer for identifying this unclear wording. The phrase “similar to carbon chains” was intended to refer to compounds with comparable carbon backbones, rather than to imply that increasing carbon-chain length enhances hygroscopicity. We have revised the sentence to clarify that, among compounds with similar carbon skeletons, greater oxygenation or the presence of additional polar functional groups generally corresponds to enhanced water uptake. We also clarify that this is a first-order tendency rather than a universal relationship, because solubility, molecular structure, phase state, and intermolecular interactions can cause substantial deviations.

[Lines 15-17] “...Structurally similar compounds may exhibit markedly different hygroscopic behavior, underscoring the role of molecular interactions; with comparable carbon number, greater oxygenation and functional groups generally correspond to higher κ_{org} ; however, molecular structure and solubility may lead to substantial deviations from this trend...”

[General Comment]

line 123 - is the (Description -- Across, n.d.) text accurate? The other campaigns have site details.

Response: We have added the ACROSS campaign and site description.

[Lines 143-145] “...One month of measurements from 16th June 2022 to 15th July 2022 during the ACROSS (Atmospheric ChemistRy Of the Suburban foreSt (Description –

Across, n.d.)) campaign in Paris was used. This “supersite” is surrounded by suburban facilities, forests, agricultural fields, and roads connecting Paris (Petit et al., 2015) ...”

[General Comment]

The names and formation of the functional groups is not consistent throughout the text, figures, and SI. Please select one way to describe them and use consistent naming, including subscript and charge indication, if included, everywhere.

Response: We have made all the functional group names uniform throughout the text and SI. They are mentioned and described at the beginning, and later, they are used everywhere as Cx, CH, CHO, and CHOgt1.

[Lines 220-225] “...One family (e.g., Cx+, CxHy+, CxHyO+, CxHyOz+, where x represents the number of C atoms, y the number of H atoms, and z > 1 the number of O-atom in the fragments) represents the sum of the organic fragments having different combinations of C, H, and O that define the family. Organic ions were grouped according to their elemental composition into the AMS ion families Cx (Cx+), CH (CxHy), CHO (CxHyO+, containing one oxygen atom), and CHOgt1 (CxHyOz+, z > 1, containing more than one oxygen atom). The abbreviation “gt1” denotes “greater than one” and follows the family notation used in the AMS data-processing workflow. These categories represent ion-fragment families rather than directly identified molecular functional groups. ...”

[General Comment]

It is not clear in the modeling sections why everything is time-dependent. How is this time component handled for the laboratory-measured parameter analysis?

Response: We thank the reviewer for identifying this ambiguity. Time is not an input variable in the laboratory-trained model. In the laboratory analysis, each observation corresponds to a specific compound and dry particle diameter, and the predictors consist of the AMS ion-family fractions, O:C, and particle diameter. The laboratory model therefore learns relationships across compounds and particle sizes rather than temporal relationships.

Time dependence applies only to the ambient application, where aerosol composition and measured hygroscopicity vary throughout each field campaign. The laboratory-trained model is applied independently to the AMS descriptors available at each ambient measurement time to produce a time-resolved κ_{org} . Time itself is not supplied as a predictor; it serves only as an observation index and, in the temporally blocked validation, as the basis for separating training and test periods. We have revised the notation and modelling description to distinguish clearly between compound–diameter-indexed laboratory observations and time–diameter-indexed ambient observations. Changes in the manuscript: Section 2.4 and Eqs. (5)-(6) were revised to distinguish the compound–diameter indexing of the laboratory dataset from the time-diameter indexing of the ambient datasets. We now clarify that time is not an input to the laboratory or ambient ML models; it is used only to index ambient observations and define temporally blocked validation periods.

[Lines 245-260] “...from 22 standard chemical compounds and ambient measurements. The laboratory and ambient stages have different data structures. In the laboratory stage, an observation is defined by compound (c) and dry particle diameter (d); no temporal variable is used as a model input. In the ambient stage, observations are indexed by measurement time (t) and particle diameter (d), because aerosol composition varies during the campaigns. Time is used only to identify observations and to construct temporally blocked validation sets; it is not included as a chemical predictor. As a general formula, specifically, the κ estimate (y_t) given some features x_t or c, d at a given time step t or compound and diameter for the laboratory stage is set to: ...”

$$\hat{y}_t \approx f_{\kappa}(x_{t \text{ or } c, d}) \dots\dots 5$$

[Lines 295-297] “...Cross-validation of the laboratory model was performed across compound-to-diameter observations and did not involve a temporal split. For laboratory validation, all diameter-resolved observations for the same compound were assigned to the same fold, preventing measurements of a given compound from appearing in both the training and validation sets...”

[General Comment]

line 251 - should the bar be on top of the y, or is there a typo in the equation above?

Response: Sorry, it was a typo; we revised it. Thank you. $\hat{y}$

[General Comment]

line 290 - it is not clear why the hygroscopicity parameter would grow with dry particle diameter. is there previous data or theory that suggests this should be the case? What previous studies does this agree with? Fig 2- Malonic only shows one literature datapoint for 200 nm.

Response: We thank the reviewer for raising this point. Under the ideal κ -Köhler framework, the effective hygroscopicity parameter of a chemically homogeneous particle is generally expected to be approximately independent of dry particle diameter after accounting for the Kelvin effect. Thus, an increase in κ with dry diameter is not an expected general theoretical behavior.

Our measurements showed an apparent increase in the growth factor and derived effective κ of malonic acid over the investigated diameter range. We have revised the manuscript so that this observation is not interpreted as an established intrinsic size dependence. Possible contributing factors may include particle phase state or morphology, incomplete equilibration, size-dependent volatilization or measurement uncertainty, and uncertainty in the Kelvin correction. The available literature value shown in Fig. 2 is for 200 nm particles only. We therefore agree that it supports agreement at that diameter but does not independently support the observed size trend. The previous statement implying agreement with earlier studies on the full-size dependence has been removed.

We have also evaluated the influence of the Kelvin correction on the diameter dependence of the derived κ values [insert the quantitative result of the sensitivity analysis here]. The remaining trend is described as apparent, and its physical origin remains unresolved.

[Refer to the R2 comment 1 response, and Fig. S10- 11 is added in the supplementary information.](#)

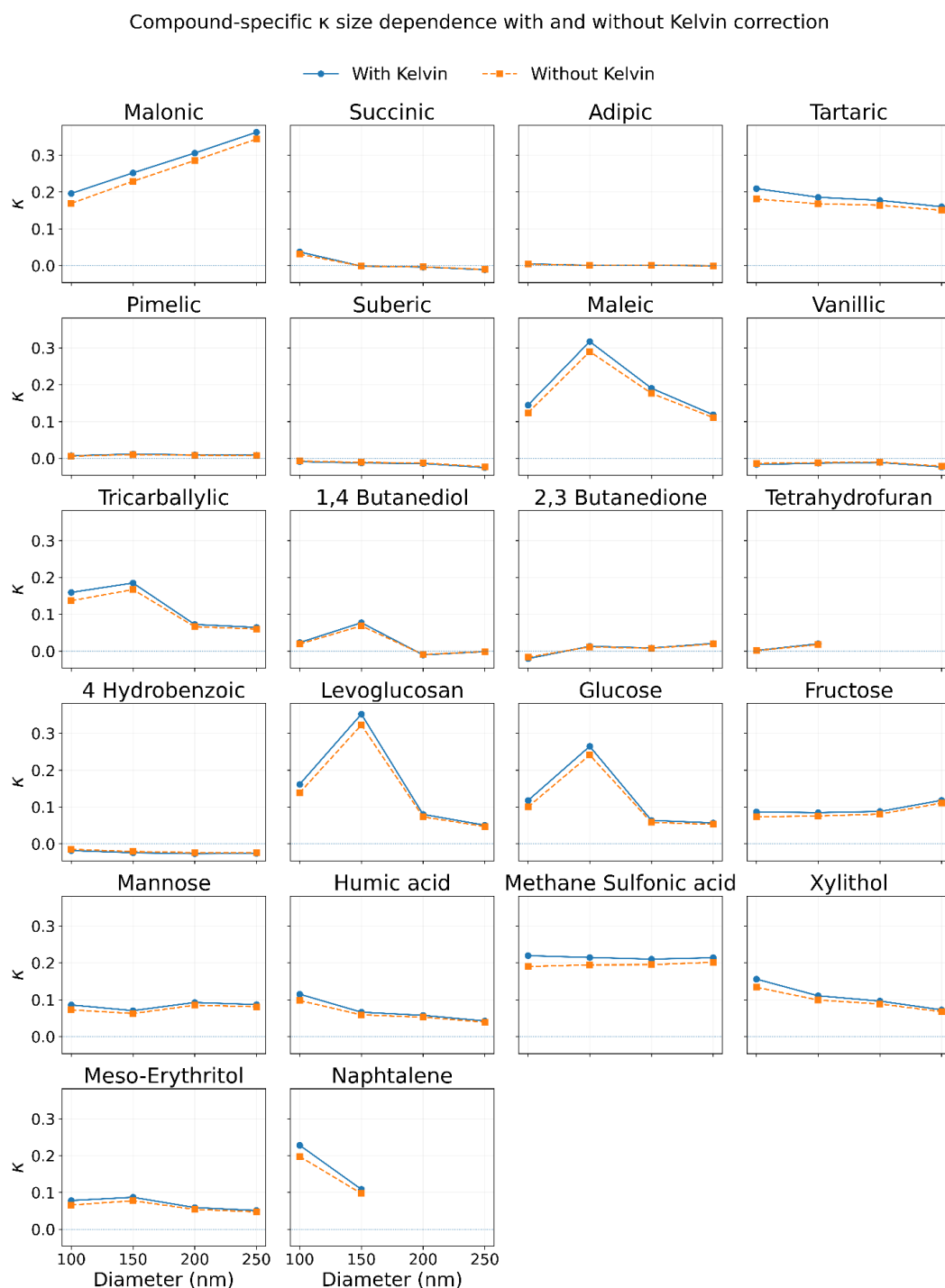


Figure S10. Sensitivity of the derived organic aerosol hygroscopicity (κ_{org}) to the Kelvin correction. Size-resolved κ_{org} for the investigated organic compounds calculated with the Kelvin correction and with the Kelvin term removed.

[General Comment]

The paragraph starting at line 285 is confusing. I think it would be helpful to have a paragraph discussing the linear dicarboxylic acids and then another paragraph discussing molecules with different functionalities, and then somehow comparing them. The way it reads now almost implies all the molecules are linear dicarboxylic acids.

Response: We thank the reviewer for this helpful suggestion. We have reorganized this section into separate paragraphs. The first paragraph now discusses linear dicarboxylic acids as a homologous class of compounds. The second discusses compounds with different carbon skeletons and functional groups, such as unsaturated or multifunctional carboxylic acids, sugars, and anhydrosugars.

[Lines 349-365] "...Compounds with different structures and functionalities were considered separately from the straight-chain dicarboxylic acids. Maleic acid, an unsaturated dicarboxylic acid, and tricarballic acid, a tricarboxylic acid, showed a sudden increase in growth at 150 nm, although their AMS ion-family fractions did not vary substantially among the investigated diameters. The physical origin of this apparent size dependence remains unresolved because comparable size-resolved literature data are unavailable. Possible explanations include size-dependent measurement effects, particle phase state or morphology, and equilibration behavior. Apparent size-dependent behavior was also observed for levoglucosan and glucose (Fig. S1), whereas fructose and mannose showed good agreement with previous measurements ((Han et al., 2022; Man et al., 2008); Figs. S1–S2 and Table S1).

Comparison among these compounds shows that carbon-chain-length effects can be evaluated most directly within the homologous straight-chain dicarboxylic-acid series. Comparisons with unsaturated, substituted, or multifunctional compounds must additionally account for differences in functional-group composition, molecular configuration, solubility, and phase state. Therefore, compounds with similar carbon numbers may exhibit substantially different water uptake when their structures and functional groups differ. Collectively, these results demonstrate that organic hygroscopicity cannot be described by a single compositional metric alone, motivating a multi-parameter framework that integrates size, functionalization, and physicochemical properties for improved parameterization of κ_{org} ..."

[General Comment]

hypothesis starting at line 300 - is there any previous work or modeling to back up this hypothesis, is is not clear why dry particle size alone would change the 'hygroscopicity of the organic compounds themselves.'

Response: We agree with the reviewer that dry particle size alone is not expected to change the intrinsic hygroscopicity of a chemically homogeneous organic compound. Under equilibrium κ -Köhler theory, κ is generally treated as approximately independent of dry diameter after accounting for the Kelvin effect. We have ruled out the hypothesis that particle size affects the hygroscopicity of the compounds themselves.

[Lines 349-365] “...Compounds with different structures and functionalities were considered separately from the straight-chain dicarboxylic acids. Maleic acid, an unsaturated dicarboxylic acid, and tricarballic acid, a tricarboxylic acid, showed a sudden increase in growth at 150 nm, although their AMS ion-family fractions did not vary substantially among the investigated diameters. The physical origin of this apparent size dependence remains unresolved because comparable size-resolved literature data are unavailable. Possible explanations include size-dependent measurement effects, particle phase state or morphology, and equilibration behavior. Apparent size-dependent behavior was also observed for levoglucosan and glucose (Fig. S1), whereas fructose and mannose showed good agreement with previous measurements ((Han et al., 2022; Man et al., 2008); Figs. S1–S2 and Table S1).

Comparison among these compounds shows that carbon-chain-length effects can be evaluated...”

[Refer to R2 comment 1 response, and Fig. S10- 11 is added in the supplementary information.](#)

[General Comment]

The authors may want to comment on the known differences in physical properties of odd and even carbon dicarboxylic acids.

Response: We thank the reviewer for this suggestion. We have added a discussion of the odd–even effect in straight-chain dicarboxylic acids. Even-carbon dicarboxylic acids commonly form more efficiently packed and stable crystal structures, which can reduce their tendency to take up water relative to neighboring odd-carbon homologues. This behavior may contribute to the contrast between malonic acid and several even-carbon dicarboxylic acids in our measurements. However, the absence of detectable growth for pimelic acid, an odd-carbon compound, shows that carbon-number parity alone does not explain the complete dataset. We therefore discuss the odd–even effect as a potentially contributing structural factor rather than a universal explanation.

[Lines 346-355] “...Although some slight variation is seen in family group fractions from AMS, and a very low change in GF (0.01-0.09) between diameters, a more or less the same family group fraction was seen for these hydrophobic dicarboxylic acids (Fig. 2). Straight-chain dicarboxylic acids are also known to exhibit odd–even alternation in their physical properties. Differences in the orientation of the terminal carboxyl groups can affect molecular packing, producing alternating trends in crystal stability, melting point, solubility, and phase behavior. Even-carbon dicarboxylic acids generally pack more efficiently in the crystalline state than neighbouring odd-carbon homologues and may consequently resist water uptake until their deliquescence conditions are reached. This effect may contribute to the contrast between malonic acid (C3) and the even-carbon succinic (C4), adipic (C6), and suberic (C8) acids. However, pimelic acid (C7) also showed little detectable growth at 90% RH, demonstrating that carbon-number parity alone cannot explain the observations...”

[Comment]

the paragraph at the beginning of section 3.2 is confusing to read. A reader is unlikely to know off hand what functional groups are in each of these molecules. It would be helpful to clarify why you are comparing certain things and how those molecules differ.

Response: We have revised the beginning of Section 3.2 to identify the principal structural differences among the compounds. Specifically, malonic and succinic acids are introduced as neighboring saturated straight-chain dicarboxylic acids; tartaric acid is distinguished by its two additional hydroxyl groups; maleic acid is identified as an unsaturated dicarboxylic acid; tricarballic acid is identified as a tricarboxylic acid; and levoglucosan is identified as an anhydrosugar containing hydroxyl and ether/acetal functionality but no carboxyl group. We also explain that these comparisons were selected to examine how carbon-backbone length, hydroxyl substitution, unsaturation, and the number and type of functional groups relate to water uptake. [Refer to revised section 3.2 from the manuscript.](#)

[Lines 380-408] "...Converting measured growth factors to the effective hygroscopicity parameter κ using Eqs. (2) and (3) enables a consistent comparison of organic hygroscopic behaviour among compounds and particle diameters within the κ -Köhler framework. Figure S3 illustrates the size-resolved κ values for selected organic compounds. These compounds were compared to examine the relationships between hygroscopicity and carbon number, as well as the presence of hydroxyl, carboxyl, and unsaturated structural features, and their dependence on molecular composition and oxidation-related ratios. The carboxylic acid group exhibits substantial variability in κ across compounds and particle diameters (Fig. S3a–e). Malonic and succinic acids are neighbouring saturated straight-chain dicarboxylic acids that differ by one carbon atom, allowing comparison within a homologous series. Malonic acid exhibits a pronounced apparent increase in κ , from approximately 0.20 at 100 nm to approximately 0.35 at 250 nm, consistent with its high water solubility. Whereas succinic acid exhibits near-zero κ across all measured sizes, indicating limited water uptake under subsaturated conditions despite its similar carbon backbone. Tartaric acid is a C4 (with four carbons) dicarboxylic acid with two additional hydroxyl groups and displayed moderate hygroscopicity with weaker apparent diameter dependence. Comparing tartaric and succinic acids therefore illustrates that compounds with the same carbon number and number of carboxyl groups can exhibit different water uptake when additional polar functional groups are present. Maleic acid is an unsaturated C4 dicarboxylic acid and showed higher effective κ at intermediate diameters but lower values at larger diameters. Its comparison with saturated C4 succinic acid highlights the potential influence of unsaturation, molecular configuration, solubility, and phase state..."

[Comment]

The references in the SI are in author (year) but there is no bibliography included with the SI and the bibliography in the main text is numbered, not alphabetical.

Response: Thank you very much for the comments. We changed and added the bibliography in the SI as well as in the main manuscript. Now it's in alphabetical order.

[Comment]

a main theme in this work is that the hygroscopicity of a single-component particle is a function of size, but it is really not clear why this should be the case. how many repeated trials were performed of the hygroscopicity experiments? what kind of validation has been done to confirm this is real and significant?

Response: We thank the reviewer for raising this point. Each selected dry diameter was measured using three repeated HTDMA scans, with each compound scan lasting approximately 60 min. The plotted values represent the mean of the three scans, and the error bars, which are already included in the figures (Fig. 2, Fig. S1-S3, and Fig. S8), represent the mean $\pm$ standard deviation of these measurements (Table S1). We have now stated this explicitly in the Methods and all relevant figure captions.

We agree that measurement repeatability alone does not demonstrate that dry particle size changes the intrinsic hygroscopicity of a single-component particle. We have therefore revised the manuscript to describe the differences among diameters as apparent size-related variations rather than as a general intrinsic size dependence. Where these differences are comparable to the replicate variability, they are not interpreted as evidence of size dependence.

[Lines 134-137] "...The physicochemical properties of the 22 studied compounds are summarised in Table 1 & S1... For each selected dry diameter, three repeated HTDMA scans were conducted, with each compound scan lasting approximately 60 min. Reported growth factors are the arithmetic means of the three scans, and measurement variability is expressed as mean $\pm$ standard deviation (n=3). These repeated scans quantify short-term measurement repeatability under the same experimental conditions..."

[Refer to the figure captions in the revised manuscript.](#)

[Comment]

How well is the model trained on data from two field campaigns likely to predict all ambient aerosol? It would be nice to see a real validation of the model using a dataset that is not used to train the model.

Response: We thank the reviewer for identifying this important limitation. We agree that validation using held-out observations from the same two campaigns does not establish that the model can predict the full diversity of ambient aerosol. The reported test results constitute internal validation within the Paris and Goldlauter datasets rather than fully independent external validation.

A separate multi-site study by us (Deshmukh et al., 2026) evaluated related hygroscopicity-estimation approaches across nine field datasets and compared site-specific and combined-site models. That study generally found that site-specific models performed as well as or better than models trained on pooled observations, indicating that relationships between aerosol properties and hygroscopicity can vary among measurement environments.

However, because that study used a different ML configuration and predictor set, we do not present its results as validation of the model developed here. Instead, they provide independent evidence supporting the need for cautious interpretation and additional cross-site testing.

A fully independent validation of the present framework would require applying the finalized model—without retraining, recalibration, feature selection, or preprocessing based on the test data—to a separate campaign containing the same required measurements. Such a harmonized dataset was not available within the scope of this study. Although we try testing between these two campaigns by holding out one side completely. This test uses the shared ambient chemical descriptors, but it does not yet reproduce the complete laboratory-to-ambient hybrid model exactly. For example, the training campaign (Paris) and the independent test campaign (Goldlauter) have 1,871 and 1,413 samples for training and testing, respectively. Shows very poor transfer with $R^2 = -4.806$. The issue may be different seasons and the lab model being excluded. We have therefore revised the manuscript to make clear that the demonstrated performance is limited to the two campaigns investigated. We have removed claims implying applicability to all ambient aerosols and explicitly identified independent multi-site validation as a necessary next step. Refer to the revised Section 3.4 and conclusion.

[Lines 562-570] “...These results highlight the importance of integrating controlled laboratory studies with data-driven ambient models to enhance the representation of aerosol hygroscopicity in atmospheric applications. The ambient-model evaluation presented here is limited to observations from the Paris and Goldlauter campaigns. Although held-out observations were used for testing, these observations originated from the same campaigns used for model development and therefore provide internal rather than fully independent external validation. The reported performance should consequently not be interpreted as demonstrating applicability to all ambient aerosols. A separate multi-site analysis using a different ML framework found that site-specific hygroscopicity models generally performed as well as or better than models trained on pooled data from multiple sites (Deshmukh et al., 2026), indicating that predictor–hygroscopicity relationships may vary among atmospheric environments. Because the model architecture and predictors differed from those used here, those results do not constitute external validation of the present model. Application to additional campaigns without retraining or recalibration is required to assess broader transferability....”

[Comment] Other comments

is there a reason the authors have chosen to reference the 2013 IPCC report rather than the newest edition?

Response: No, there was no specific reason this IPCC report was the most cited; hence, it was used. We also added the new IPCC 2021 report in the introduction.

[Lines 30-31] “...Atmospheric aerosols significantly impact the Earth's radiation budget, exerting both direct and indirect effects (IPCC, 2013; Rosenfeld et al., 2014; Li et al., 2016; (IPCC, 2021)) ...”

[Comment]

please check that abbreviations are defined at the first instance and then used throughout the text, there are instances of the same abbreviation being defined multiple times.

Response: We revised the sentence and added κ and other abbreviations at their first occurrence. Thank you.

Response to Referee comments #RC2

1. [General Comment] The manuscript “From Single Compounds to Ambient Aerosols: A Machine-Learning-Based Estimation of Organic Hygroscopicity” presents a machine-learning (ML) framework for predicting aerosol hygroscopicity using laboratory hygroscopic tandem differential mobility analyzer (HTDMA) measurements combined with aerosol mass spectrometry (AMS) chemical characterization. The work falls within the scope of Atmospheric Chemistry and Physics, addresses a scientifically relevant topic, and proposes a novel approach for predicting organic aerosol hygroscopicity from chemical descriptors. The manuscript demonstrates that the ML framework improves the prediction of ambient aerosol hygroscopicity compared with Zdanovskii–Stokes–Robinson and Monte Carlo approaches.

However, the practical application of the proposed framework remains somewhat unclear, and several scientific and methodological issues should be addressed before publication.

Response: We thank the reviewer for the constructive comment and positive summary of our work. Below, we list a point-by-point response to all questions and suggestions that improved the paper.

[Major Comment] The manuscript repeatedly highlights a size dependence in organic aerosol hygroscopicity. However, the physical origin of this dependence is insufficiently discussed. Could the observed size dependence be related to Kelvin-effect corrections used in the derivation of κ from HTDMA measurements? In that case, the effective water activity at the particle surface may differ from the ambient relative humidity. However, this explanation appears unlikely because the growth factor does not exhibit a consistent dependence on particle size across different compounds. Some compounds show increasing hygroscopicity with size, others decreasing hygroscopicity, and some show little dependence at all.

Response: We thank the reviewer for pointing out the need for a more detailed discussion of the observed size dependence in organic aerosol hygroscopicity. We agree that the Kelvin contribution in the HTDMA-derived κ should be explicitly considered because the conversion from growth factor to κ depends on the wet particle diameter.

We therefore performed an additional sensitivity analysis in which κ was recalculated with and without the Kelvin contribution. The results show that the Kelvin term introduces a systematic, diameter-dependent contribution to the inferred κ , particularly for smaller

particle sizes. However, the Kelvin contribution alone does not explain the observed size dependence across the full compound set. In particular, the experimentally derived κ values exhibit compound-specific behavior, with some compounds showing increasing κ with particle size, others decreasing κ , and others showing little dependence. These different trends cannot be explained by a single systematic Kelvin-related effect.

In the original analysis, the Kelvin term was already included in the conversion of HTDMA growth factors to κ :

$$\kappa_{measured} = (GF^3 - 1) \left(\frac{\exp \left(\frac{A}{D_{dry} \cdot GF} \right)}{RH} - 1 \right)$$

To assess its influence on the observed size dependence, we performed an additional sensitivity analysis by recalculating κ from the same measured GF values after removing the Kelvin term, i.e. using

$$\kappa_{noKelvin} = (GF^3 - 1) \left(\frac{1}{RH} - 1 \right)$$

We have therefore revised the manuscript to distinguish between the mathematical size dependence introduced during the HTDMA κ inversion and the remaining experimentally observed, compound-specific size dependence. However, comparison of the κ -diameter slopes with and without the Kelvin contribution shows that the compound-specific trends are largely preserved: 16 of 22 compounds exhibit decreasing κ with size when the Kelvin correction is included, compared with 15 of 22 without it, while 6 and 7 compounds, respectively, show increasing behaviour (Fig. S10). We now discuss possible contributions from compound-dependent particle phase state, dissolution behaviour, and other differences in water-organic interactions. We emphasise that the present measurements do not allow us to uniquely identify a single mechanism responsible for the observed trends.

[Lines 350-360] The following statement has been added to address these knowledge gaps; “...The observed variation of κ_{org} with particle diameter requires careful consideration because κ is derived from HTDMA growth factors using a Köhler framework that includes a Kelvin contribution. Since the Kelvin term depends on wet particle diameter, part of the apparent size dependence of κ may arise from the GF-to- κ conversion rather than from differences in water uptake alone. We therefore evaluated the sensitivity of the derived κ values to the Kelvin correction (Fig. S10-11). The Kelvin contribution introduces a systematic diameter-dependent effect, particularly at smaller particle sizes, but does not account for the range of compound-specific trends observed in the measurements. Specifically, both increases and decreases in κ_{org} with particle size, as well as weak or negligible size dependence, are observed. The persistence of these trends after removing the Kelvin contribution indicates that the observed variability cannot be attributed solely to the Kelvin correction. Possible contributors include differences in particle phase state, molecular packing or crystallinity, dissolution and equilibration behaviour, and compound-specific water-organic interactions. As these processes were not independently constrained in the present experiments, we do not assign a unique physical mechanism to the observed

size dependence. Although functional-group classifications capture first-order trends, substantial variability remains even among structurally similar compounds. Together, these results suggest that single compositional descriptors such as O:C or CO₂/CH alone are insufficient to describe the observed variability in κ_{org} , motivating the use of a multi-parameter, physically constrained machine-learning framework....”

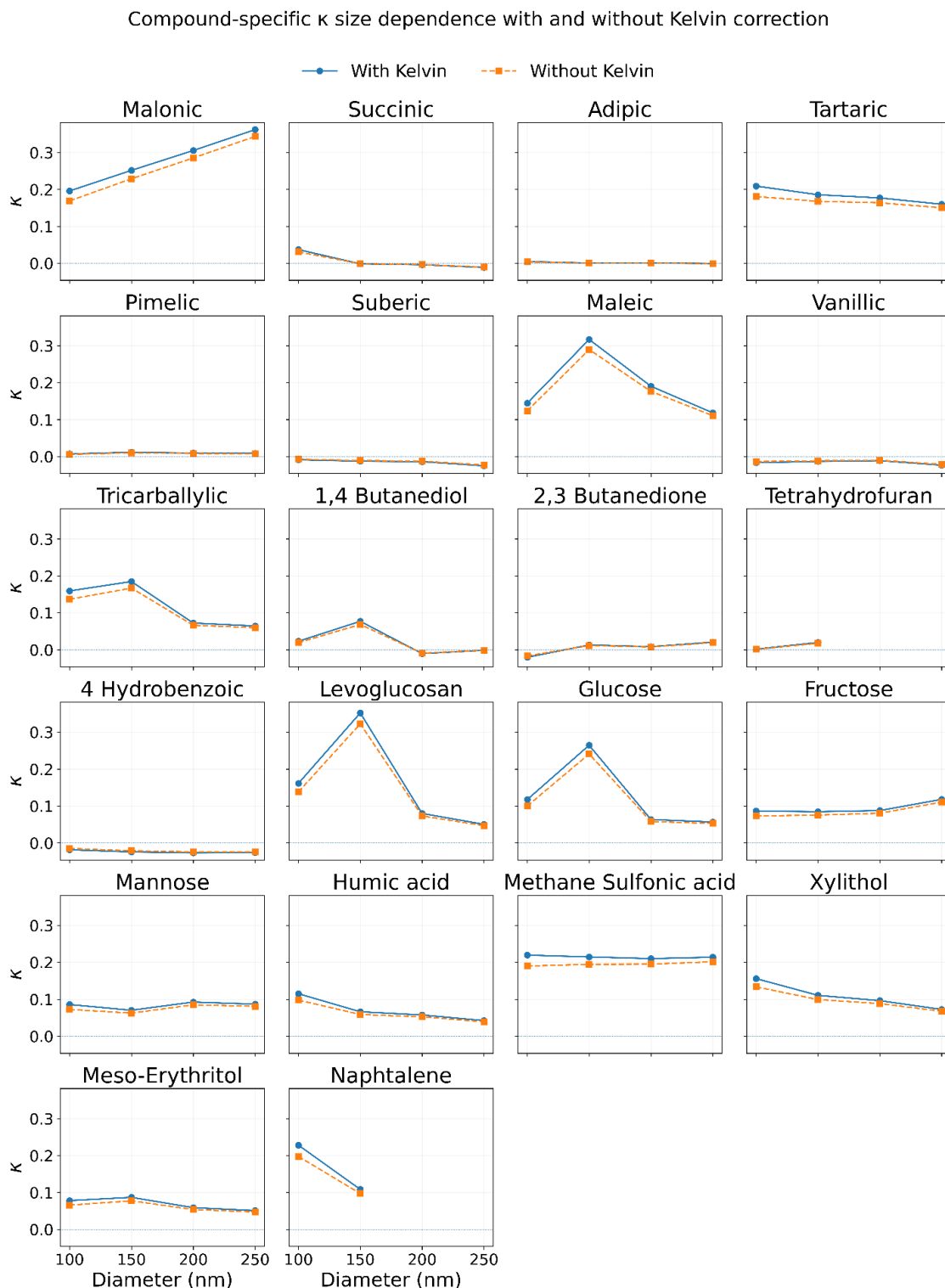


Figure S10. Sensitivity of the derived organic aerosol hygroscopicity (κ_{org}) to the Kelvin correction. Size-resolved κ_{org} for the investigated organic compounds calculated with the Kelvin correction and with the Kelvin term removed.

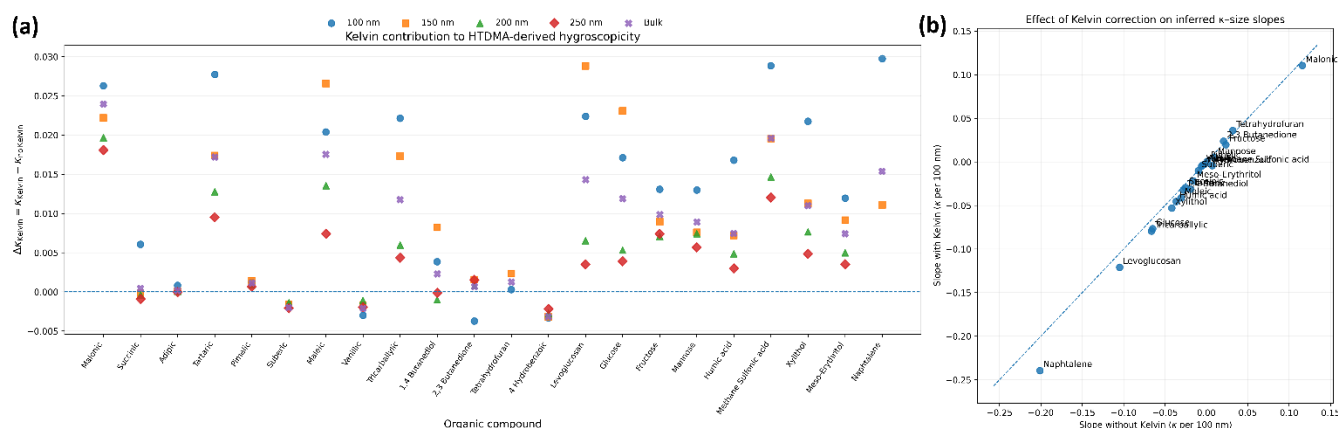


Figure S11. Sensitivity of the derived organic aerosol hygroscopicity (κ_{org}) to the Kelvin correction. (a) Mean and median difference between κ calculated with and without the Kelvin contribution as a function of dry particle diameter. (b) Linear κ_{org} –diameter slopes for each compound obtained with and without the Kelvin correction. The comparison illustrates the contribution of the Kelvin correction to the inferred magnitude of size dependence and whether the direction of compound-specific trends is preserved.

[Major Comment] The Introduction omits discussion of established thermodynamic frameworks used for predicting hygroscopic growth in inorganic-organic aerosol mixtures, such as AIOMFAC, E-AIM, and UManSysProp. Why is an ML approach needed? What limitations of existing thermodynamic models does it address? Does the proposed framework improve predictive skill, computational efficiency, applicability, or interpretability? At minimum, the authors should discuss these existing approaches in the Introduction and clearly position their work relative to them. Ideally, some comparison against one or more thermodynamic models should be included.

Response: We thank the reviewer for highlighting this and for the need to better position the proposed machine-learning framework relative to, and to clarify, its potential implementation in atmospheric and climate models. We agree that thermodynamic frameworks such as AIOMFAC, E-AIM, and UManSysProp are important established approaches for describing aerosol water uptake and inorganic–organic interactions. We have therefore expanded the Introduction to discuss these approaches and to clarify the complementary role of the present ML framework.

Thermodynamic models provide a physically based description of aerosol water uptake and phase behaviour and are particularly powerful when detailed chemical composition, molecular structures, functional-group information, and thermodynamic properties are available. For example, AIOMFAC represents organic compounds using functional groups and explicitly accounts for organic–inorganic interactions, while E-AIM calculates equilibrium gas/liquid/solid partitioning, aerosol water content, and so on.[Zuend et al., 2011; Wexler and Clegg, 2002; Topping et al., 2016].

The motivation for the present ML framework is therefore not to replace these thermodynamic models, but to address a complementary problem: predicting effective aerosol hygroscopicity from routinely available bulk aerosol measurements, for which

molecular-level chemical speciation is generally unavailable. The present framework uses AMS-derived chemical-family information and O:C descriptors, both of which can be obtained from large observational datasets. This provides an empirical mapping between readily measurable aerosol chemical characteristics and the effective hygroscopicity inferred from HTDMA observations, without requiring explicit specification of the complete molecular composition and thermodynamic parameters of the organic mixture.

In the present study, the ML framework substantially improves predictive skill relative to the bulk-chemistry/ZSR-based approach evaluated here. However, because the observational dataset does not contain the detailed molecular composition required for a consistent AIOMFAC/E-AIM calculation for all samples, we have avoided interpreting this result as evidence that the ML framework is intrinsically more accurate than established thermodynamic models. Instead, we now explicitly describe the two approaches as complementary: thermodynamic models provide physically based predictions when sufficiently detailed chemical information is available, whereas the present ML framework is intended for observationally constrained prediction of effective hygroscopicity from bulk aerosol descriptors.

Regarding implementation in large-scale atmospheric models, we agree that the present framework should not be described as directly plug-and-play, because global models generally do not simulate AMS ion/family variables in the same form as the observations used here. We have therefore revised the relevant statements in the manuscript to clarify that implementation would require a mapping between model-resolved aerosol properties and the chemical descriptors used by the ML framework.

The intended implementation of developing and evaluating this translation for a particular global model is beyond the scope of the present study and would require model-specific testing. We have revised the manuscript accordingly and have moderated our previous statements regarding direct implementation in climate models.

[Lines 75-92] The following statement has been added to address these knowledge gaps; “...Established thermodynamic frameworks provide physically based descriptions of aerosol water uptake and phase behaviour in inorganic–organic mixtures. The Extended Aerosol Inorganics Model (E-AIM) calculates equilibrium gas–liquid–solid partitioning, aerosol water content, and phase state for specified chemical systems(Wexler and Clegg, 2002) , while AIOMFAC represents non-ideal interactions among inorganic ions, water, and organic functional groups through a group-contribution activity-coefficient framework (Zuend et al., 2011) . UManSysProp provides complementary molecular-structure- and group-contribution-based tools for predicting organic aerosol properties and related hygroscopic growth and cloud-activation behaviour(Topping et al., 2016) . These approaches are particularly valuable when the molecular identities or functional-group compositions of the aerosol constituents, as well as the required thermodynamic parameters, are available. However, ambient organic aerosol is generally not resolved at this level of molecular detail by bulk online measurements, and such information is also not routinely represented in large-scale atmospheric models. The machine-learning framework developed here, therefore, addresses a complementary input-data regime by relating effective aerosol hygroscopicity to AMS-derived ion-family fractions, oxidation-

related descriptors, and bulk aerosol composition available from field observations. It does not explicitly calculate activity coefficients or phase equilibria and is not intended to replace thermodynamic models. Instead, it provides an observationally constrained parameterization of effective hygroscopicity when complete molecular speciation of the organic mixture is unavailable. Its implementation in atmospheric models would require a model-specific mapping from simulated aerosol composition and oxidation state to descriptors consistent with the ML training...”

[Major Comment] The manuscript repeatedly claims that the framework can improve the representation of aerosol hygroscopicity in atmospheric and climate models (e.g. lines 25, 102, and 496). However, no concrete pathway for incorporating the framework into such models is presented. The ML framework relies on AMS-derived family-group information and O:C ratios, whereas global climate models typically do not simulate aerosol composition at this level of chemical detail. The authors should clarify, how the proposed framework could realistically be implemented in large-scale atmospheric models. Which model variables would be used as inputs?

Response: We appreciate the insightful observation regarding the practical implementation of our framework. We agree that the previous manuscript did not sufficiently clarify how the AMS-derived descriptors used here could translate to large-scale atmospheric models. We have therefore revised the manuscript to clarify that the present framework is an observationally constrained parameterization rather than a directly plug-in climate-model scheme.

In large-scale models, the ML framework could be implemented by mapping model-resolved aerosol properties, such as organic aerosol mass, O:C or other oxidation/aging indicators, and inorganic aerosol species (e.g., sulfate, nitrate and ammonium), to the chemical descriptors used by the ML model. The predicted effective κ could then be used within the model's existing aerosol water-uptake and cloud-activation calculations. Because the detailed AMS family-group information is generally not explicitly simulated by global models, developing this mapping for a specific model would require additional model-specific evaluation and is beyond the scope of the present study. We have accordingly clarified this pathway and revised our previous statements regarding direct climate model implementation.

[Lines 24-25] “...By bridging controlled laboratory measurements with data-driven ambient analysis, this study provides an observational basis for developing composition-dependent parameterizations of organic aerosol hygroscopicity, with potential relevance to aerosol water uptake and radiative effects...”

[Lines 119-120] “...This hybrid methodology provides a pathway for to representing κ_{org} variability in observational datasets and establishes a basis for the future development of model-compatible hygroscopicity parameterizations...”

[Lines 602-615] The following statement has been added to address these knowledge gaps: “...The present framework is not directly applicable to most large-scale atmospheric models because it relies on AMS-derived ion-family fractions, which are generally not

simulated. A model-compatible implementation would therefore require the development of a reduced input set, such as organic aerosol mass fraction, dry particle size or aerosol mode, inorganic composition (e.g., sulfate, nitrate, and ammonium), and O:C. The resulting effective κ_{org} could be combined with model-resolved inorganic components within the model's existing aerosol water-uptake treatment. Alternatively, models with sufficiently detailed organic aerosol tracers could develop an offline mapping from their source-volatility- or ageing-resolved representation to descriptors consistent with the ML training domain. Both approaches would require model-specific retraining, an assessment of applicability domains, and independent evaluation. Thus, the present study provides an observational basis for future parameterization development rather than a directly transferable climate model scheme. An application to cloud activation would require additional validation, as the framework was developed using subsaturated HTDMA measurements.”

[Major Comment] In the conclusion (line 495), the authors state: “Together, this work provides a scalable, physically grounded pathway from molecular-scale organic composition to predictions of aerosol growth, cloud interactions, ...” This statement appears too strong given the evidence presented. The framework is trained and evaluated using HTDMA-derived hygroscopicity under subsaturated conditions. No cloud condensation nuclei (CCN) measurements are included. It is therefore unclear whether the predicted κ values can be directly applied to cloud activation studies. Previous studies show difference in hygroscopicities in sub-saturated conditions and those based on CCN measurements (e.g., Petters, M. D. and Kreidenweis, S. M. (2007), A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmos. Chem. Phys.*, 7, 1961-1971.)

Response: We thank the reviewer for pointing this out. We agree that the original statement overstated the scope of the present study. Our framework is trained and evaluated using HTDMA-derived hygroscopicity under subsaturated conditions, and no CCN measurements are included. We have therefore removed statements implying that the present framework directly predicts CCN activity, cloud interactions, or hygroscopicity under supersaturated conditions. The Conclusion now explicitly limits the demonstrated applicability of the framework to subsaturated aerosol water uptake and states that extension to cloud activation would require independent evaluation against CCN measurements. We have also revised related statements elsewhere in the manuscript to maintain consistency.

[Lines 315-320] “...This framework estimates the hygroscopicity of subsaturated aerosols over the investigated particle-size range of 100–250 nm, which directly controls aerosol optical properties and size distributions and, hence, aerosol-radiation interactions (Shen et al., 2021). Its applicability to optical effects, radiative forcing, or supersaturated cloud activation requires separate evaluation...”

[Lines 605-615] The following statement has been added to address these; “...The present framework was developed and evaluated using HTDMA measurements at 90% RH and

therefore applies specifically to aerosol water uptake under subsaturated conditions. Although the κ -Köhler framework provides a conceptual connection between hygroscopic growth and cloud-droplet activation, an effective κ derived from subsaturated measurements cannot necessarily be applied directly under supersaturated conditions. Differences may arise from compound solubility, phase state, non-ideal solution behaviour, surface-tension effects, and water-uptake kinetics (Petters and Kreidenweis, 2007). Extending the present approach to CCN activity would therefore require independent validation against size-resolved CCN measurements.”

[Specific Comment] Line 172: I do not understand the following sentence: “Changing κ_i in calculations doesn’t significantly change κ_{chem} prediction, as lower soluble compounds like organics generally show only slight hygroscopic growth at relative humidities less than 98% (Petters et al., 2009; Wex et al., 2009)”. Please clarify the reasoning. The statement appears inconsistent with the demonstrated importance of organic hygroscopicity throughout the manuscript.

Response: We agree that the original sentence was unclear and could be interpreted as implying that organic hygroscopicity is unimportant. We have therefore removed the original reasoning and revised the text to more clearly describe the role of κ values in the chemical closure calculation.

[Lines 195-200] The following statement has been added to address these knowledge gaps; “...Accordingly, variability in κ_{org} can substantially affect the calculated bulk hygroscopicity when organics constitute a large fraction of the aerosol volume. The constant value $\kappa_{org} = 0.1$ (Table 2) used in the conventional ZSR calculation is therefore treated here as a reference assumption rather than a universally representative value (Petters et al., 2009; Wex et al., 2009).”

[Specific Comment] Line 231: Monte Carlo is not a machine-learning model. The sentence should be revised accordingly. There is also a typo in the beginning of the sentence. “Four-machine learning models” should be “Four machine learning models”

Response: We thank the reviewer for identifying this terminology issue. We agree that Monte Carlo sampling is not itself a machine-learning model. We have corrected the sentence to distinguish the four machine-learning models from the Monte Carlo procedure used for uncertainty estimation/parameter sampling. We have also corrected the typo from “Four-machine learning models” to “Four machine learning models.”

[Lines 270-275] The following statement has been added to address these knowledge gaps; “...Three supervised regression approaches were evaluated for estimating hygroscopicity: Multiple Linear Regression (MLR), Random Forest, and XGBoost. In addition, Monte Carlo sampling and bootstrapping were used to quantify parameter and prediction uncertainty; these procedures are not treated as machine-learning models. The parameterization of (Raparathi et al., 2025) was evaluated separately as a literature-based

comparison. Among the supervised models, Random Forest and XGBoost generally showed the strongest predictive performance. Four models are tested and run to estimate hygroscopicity...”

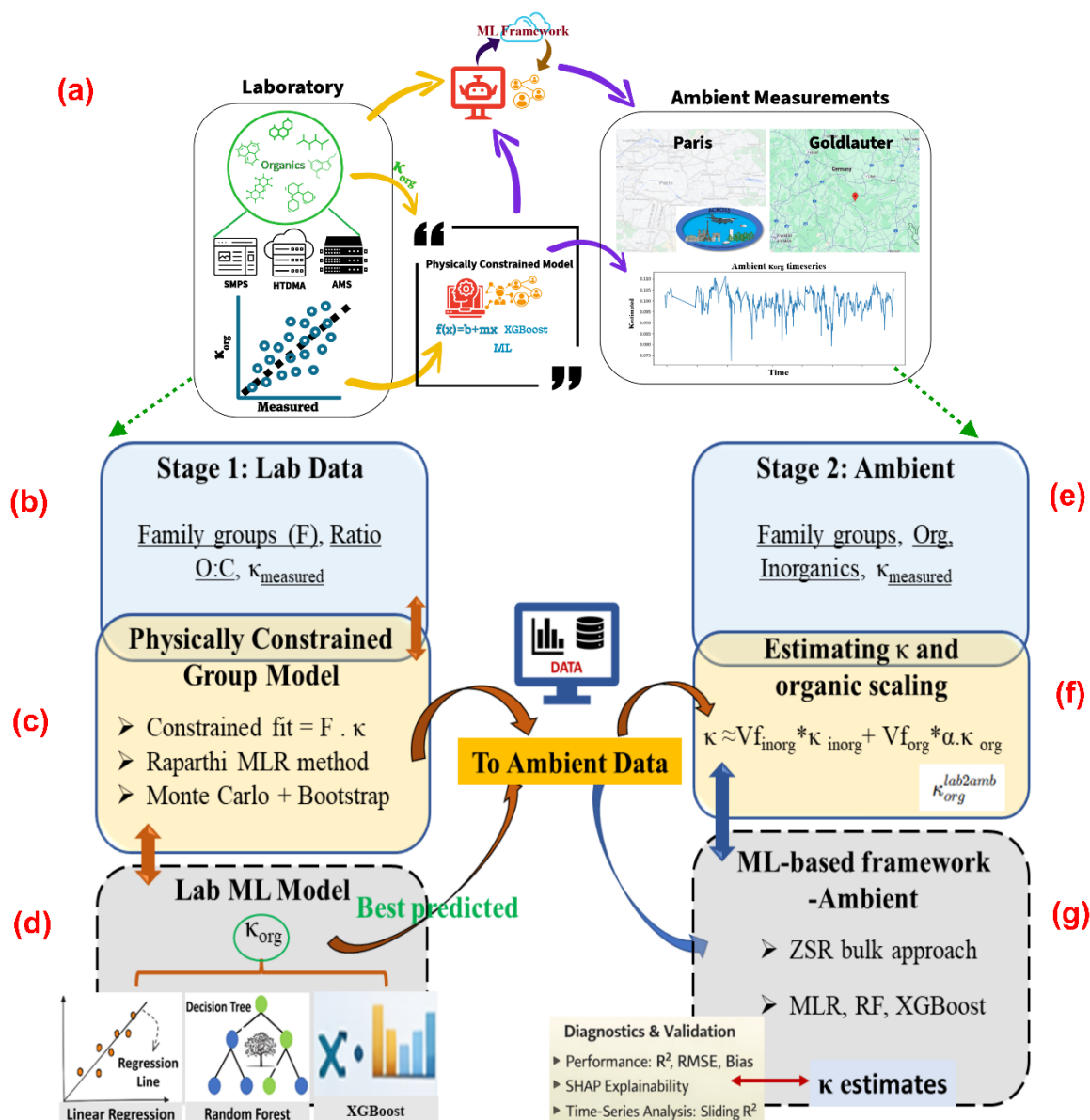
[Lines 65] “...Figure 3: Comparison of measured size-resolved organic hygroscopicity (κ_{org}) for laboratory experiments with estimates derived from (a) Monte Carlo-based constrained regression, (b) the literature parameterization of (Raparthi et al., 2025) MLR, and two (c) Random Forest & (d) XGboost machine-learning models...”

[Technical Comment] The manuscript would benefit from a thorough language check. For example: Missing articles (“a”, “an”, “the”). Inconsistent verb tenses. Occasional grammatical errors. Several overly long and difficult-to-follow sentences.

Response: The manuscript has been thoroughly revised for language and readability. We corrected missing and incorrect articles, standardized verb tense and terminology, resolved grammatical errors, and divided several long sentences into shorter and clearer statements.

[Technical Comment] Figure 1: The figure contains text that is difficult to read (e.g. “Decision Tree (1)” and “Regression time”). Because the figure is a bitmap rather than vector graphics, readability does not improve when zooming. I recommend increasing font sizes and providing a higher-resolution or vector version of the figure.

Response: We appreciate the insightful observation. We changed the figure, and we will submit the vector version/high-res plots.



[Recommendation] The manuscript addresses an important topic and presents an interesting and potentially useful ML-based framework. However, the scientific motivation and practical applicability of the approach require further clarification, particularly with respect to existing thermodynamic models, the physical interpretation of the observed size dependence, and the pathway toward atmospheric-model applications. I therefore recommend major revisions.

Response: We appreciate the insightful observation regarding the scope of our contribution and the opportunity to clarify the significance of this work. We therefore revised the manuscript and added further analysis (both in the supplementary material and in the main text). This improves our manuscript, and we thank you for the comments.

Deshmukh, S., Ferrer-Cid, P., Romshoo, B., Poulain, L., Barcelo-Ordinas, J. M., Garcia-Vidal, J., Christodoulou, A., Bezantakos, S., Denjean, C., D'Anna, B., Formenti, P., Mukherjee, S., Habib, G., Kumar, P., Huang, S., Wu, Z., Wehner, B., Henning, S., Viana, M., Petters, M. D., Ahlawat, A., and Pöhlker, M.: Regional aerosol hygroscopicity influences radiative forcing globally, *Communications Earth & Environment* 2026 7:1, 7, 416-, <https://doi.org/10.1038/s43247-026-03505-z>, 2026.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmos. Chem. Phys.*, 7, 1961–1971, <https://doi.org/10.5194/ACP-7-1961-2007>, 2007.

Raparthi, N., Wexler, A. S., and Dillner, A. M.: Hygroscopicity of Organic Compounds as a Function of Their Physicochemical Properties, *ACS ES&T Air*, <https://doi.org/10.1021/ACSESTAIR.5C00061>, 2025.

Topping, D., Barley, M., Bane, M. K., Higham, N., Aumont, B., Dingle, N., and McFiggans, G.: UManSysProp v1.0: An online and open-source facility for molecular property prediction and atmospheric aerosol calculations, *Geosci. Model Dev.*, 9, 899–914, <https://doi.org/10.5194/GMD-9-899-2016>, 2016.

Wexler, A. S. and Clegg, S. L.: Atmospheric aerosol models for systems including the ions H^+ , NH_4^+ , Na^+ , SO_4^{2-} , NO_3^- , Cl^- , Br^- , and H_2O , *Journal of Geophysical Research Atmospheres*, 107, ACH 14-1, <https://doi.org/10.1029/2001JD000451>;WGROUP:STRING:PUBLICATION, 2002.

Zuend, A., Marcolli, C., Booth, A. M., Lienhard, D. M., Soonsin, V., Krieger, U. K., Topping, D. O., McFiggans, G., Peter, T., and Seinfeld, J. H.: New and extended parameterization of the thermodynamic model AIOMFAC: Calculation of activity coefficients for organic-inorganic mixtures containing carboxyl, hydroxyl, carbonyl, ether, ester, alkenyl, alkyl, and aromatic functional groups, *Atmos. Chem. Phys.*, 11, 9155–9206, <https://doi.org/10.5194/ACP-11-9155-2011>, 2011.

AR6 Climate Change 2021: The Physical Science Basis | Climate Change 2021: The Physical Science Basis, <https://www.ipcc.ch/report/ar6/wg1/>, last access: 31 August 2024., 2024.