communications earth & environment

Article

A Nature Portfolio journal

https://doi.org/10.1038/s43247-026-03505-z

Regional aerosol hygroscopicity influences radiative forcing globally

Shravan Deshmukh[1] ✉, Pau Ferrer-Cid[2], Baseerat Romshoo[1], Laurent Poulain[1], Jose M. Barcelo-Ordinas[2], Jorge Garcia-Vidal[2], Aliki Christodoulou[3,19], Spyros Bezantakos[3], Cyrielle Denjean[4], Barbara D'Anna[5], Paola Formenti[6], Subrata Mukherjee[7], Gazala Habib[8], Prashant Kumar[9,10], Shan Huang[11], Zhijun Wu[12], Birgit Wehner[1], Silvia Henning[1], Mar Viana[13,14], Markus D. Petters[15,16], Ajit Ahlawat[1,17] ✉ & Mira Pöhlker[1,18] ✉

Aerosol hygroscopicity is a critical parameter for predicting radiative forcing and climate sensitivity, particularly under sub-saturated regimes where it drives complex aerosol–water interactions. Here, we show that externally mixed aerosols exert a stronger influence on direct radiative forcing than is currently represented in models. Incorporating our findings into radiative forcing calculations indicates a stronger aerosol cooling effect, especially at suburban sites, highlighting the importance of representing regional differences in mixing state. The conventional bulk-chemistry approach, which assumes volume-based mixing with limited spatial variability, exhibits low predictive performance for aerosol hygroscopicity ($R^2 \approx 0.61$) at urban and suburban sites. Using an interpretable machine learning framework trained on geographically diverse, region-specific datasets can capture this variability with higher accuracy ($R^2 \approx 0.97$), identifying key chemical compositional and mixing-state drivers.

Atmospheric aerosols significantly influence the Earth's radiation balance, exerting both direct and indirect effects at local, regional, and global scales. Aerosols scatter and absorb solar radiation, exerting direct radiative effects[1]. They also indirectly influence the climate by acting as cloud condensation nuclei (CCN), affecting cloud microphysics and their lifetime[1,2]. Aerosol hygroscopicity (κ) is one of the key parameters used in radiative forcing calculations and influences the uncertainties in climate predictions[3]. Sub-saturated hygroscopicity is classically described by the growth factor (GF)[4], which is most commonly directly measured by a hygroscopicity tandem differential mobility analyser (HTDMA) instrument[5]. The GF is then converted into κ[6] providing a similar proxy to CCN activity measurements. Although CCN activity has been extensively studied[7], aerosol hygroscopic growth under sub-saturated conditions remains poorly characterised, especially in remote and pristine regions[8,9]. Thus, this highlights the need for investigations into sub-saturated hygroscopicity, specifically considering different aerosol mixing states.

The mixing state of atmospheric particles is complex: internal mixtures have uniform composition and predictable hygroscopic growth[10], while external mixtures exhibit particle-to-particle chemical variability, reflecting diverse sources and atmospheric processing history. The distinction between these mixing states is crucial, as it governs the water uptake[8,11,12] of the individual particle and aerosol-radiation estimations, particularly for near-emission sources[13,14]. However, due to sparse measurements, existing estimates of hygroscopicity often rely on simplified models, such as the Zdanovskii–Stokes–Robinson (ZSR) mixing rule and assuming internal mixing[12]. Climate models often use this mixing rule, considering a certain fixed κ value for each aerosol component without accounting for regional variability in aerosol composition, particularly in the organic fraction, and the possibility of external mixing[15]. As a result, the ZSR mixing rule may oversimplify the complex relationship between hygroscopic growth and factors such as aerosol chemical composition, mixing state, and ambient relative humidity (RH), which strongly modulates water uptake, leading to uncertainties in climate simulations[8,12]. The influence of κ on aerosol optical properties and radiative forcing is minimal under dry conditions but increases sharply above ~70–80% RH, when water uptake significantly enhances scattering[16].

It remains a challenge to accurately represent variations in mixing state and regional differences in hygroscopicity, and to understand how their misrepresentation may affect the interpretation of aerosol physicochemical properties. Urban and suburban environments predominantly feature externally mixed aerosols due to fresh primary emissions from traffic and industries, superimposed on a hygroscopic background of aged secondary aerosols[17]. In contrast, marine environments are characterised by highly hygroscopic aerosols, primarily composed of sea spray with sulfate and chloride-rich secondary particles[18]. Recognising these spatial and compositional variations in aerosol mixing state and hygroscopicity is essential for

A full list of affiliations appears at the end of the paper. ✉e-mail: deshmukh@tropos.de; A.S.Ahlawat@tudelft.nl; poehlker@tropos.de

improving aerosol direct radiative effects estimations, constraining uncertainties in climate models, and advancing air quality assessments[19]. Overall, the currently used simplistic approach to computing κ is insufficient to capture these complex variations in aerosol mixing state.

Here, we address these knowledge gaps by developing an explainable machine learning (ML) framework to estimate size-segregated κ across diverse atmospheric environments (Fig. 1). The framework integrates multi-site observations spanning ten locations and multiple particle sizes (50-300 nm). Unlike earlier regional ML studies[20], our approach is extended and evaluated across geographically diverse and regionally resolved datasets, improving predictive accuracy and interpretability. By integrating chemical composition, particle number-size distribution, and meteorology, it captures the complexity of aerosol mixing states (internal and external) while enabling the imputation of data gaps. Model interpretability is explicitly assessed using SHapley Additive exPlanations (SHAP) analysis[21] to identify chemically and physically meaningful drivers influencing κ variability. We intended this framework as a regionally or site-specific, interpretable modeling approach rather than a fully generalizable, transferable model. Further examined where and why predictive skill declines across regions, thereby identifying current limitations and key priorities for future research. Finally, by coupling ML-predicted hygroscopicity (κ_{ML}) with radiative forcing calculations, we demonstrate how region-specific aerosol representations can reduce uncertainties in estimating aerosol–radiation interactions in climate models.

Results and discussion

Aerosol hygroscopicity across geographically diverse environments

A significant effect of externally mixed particles on κ was observed, particularly in urban and suburban regions where fresh emissions interact with aged aerosols. In such regions, where external mixing dominates, conventional bulk chemistry-based models (κ_{chem}) exhibit the largest errors because they assume ideal internal mixing and overlook size and source variability. The ML-derived hygroscopicity (κ_{ML}) compensates for this limitation and closely reproduces measured values ($R^2 \approx 0.97$, slope = 1.05; Fig. 2B), compared with κ_{chem} ($R^2 \approx 0.61$, slope = 0.73; Fig. 2A).

The κ exhibits substantial spatial and temporal variability across the ten sites, reflecting differences in aerosol composition, emission patterns, long-range transport, and aging processes (see Supplementary Note 2). This variability of κ from urban, remote, open-ocean, and mixed-influence environments over different time periods is evident in Figs. S2 and S3. Across these environments, κ_{ML} consistently improves agreement with measurements, although predictive skill varies with aerosol regime. It is highest at urban and suburban sites ($R^2 \approx 0.85$-0.95), moderate at marine and remote locations ($R^2 \approx 0.6$-0.85), and lowest in highly polluted environments ($R^2 \approx 0.48$), where emission sources and rapidly evolving aerosol mixtures increase hygroscopic variability. In regions dominated by complex organic-inorganic or marine aerosols, κ_{chem} underestimates hygroscopicity because important contributors, such as sea spray components or

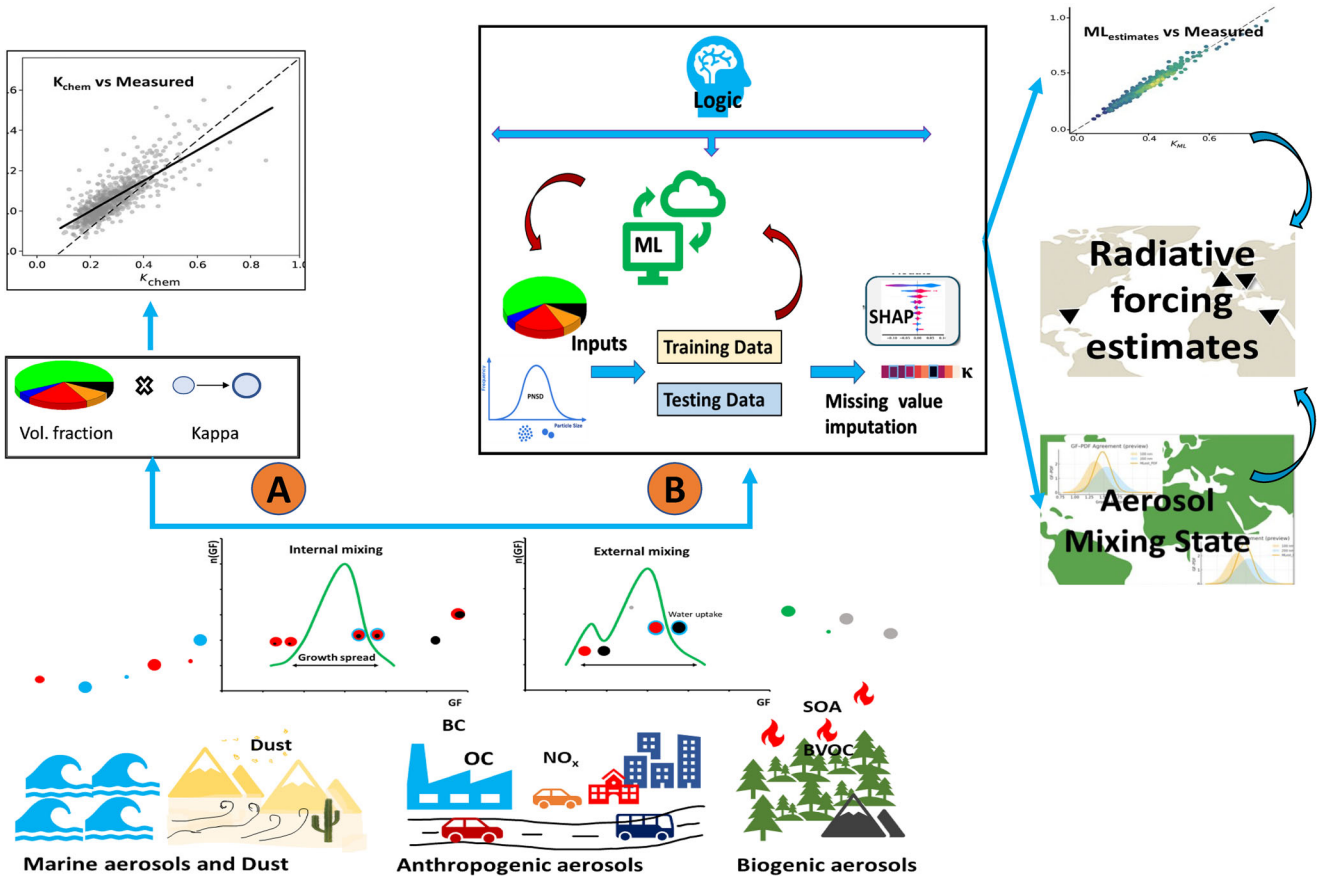

Fig. 1 | Illustrative representation of the framework used for the study with the classical ZSR and ML methods. A diverse dataset of atmospheric environments, ranging from marine to anthropogenic, was used for the following estimations. A simplified empirical classical model (A), such as the Zdanovskii–Stokes–Robinson (ZSR) mixing rule, which uses individual volume fractions and κ of chemical species (i.e., Organics and inorganics) and black carbon (BC), solely assumes internal mixing due to the bulk chemical composition of aerosols. On the other hand, (B) our interpretable data-driven approach that leverages an explainable machine learning (ML) framework integrates multi-site observations, a regional site-specific model using chemical composition, particle number size distribution (PNSD), and meteorology, additionally offering interpretability (SHAP). Incorporating κ_{ML} in first-order radiative forcing calculations highlights the importance of region-specific aerosol representations and the mixing state of particles. Furthermore, transferability and generalization are out of the scope due to the data limitations and regional variability.

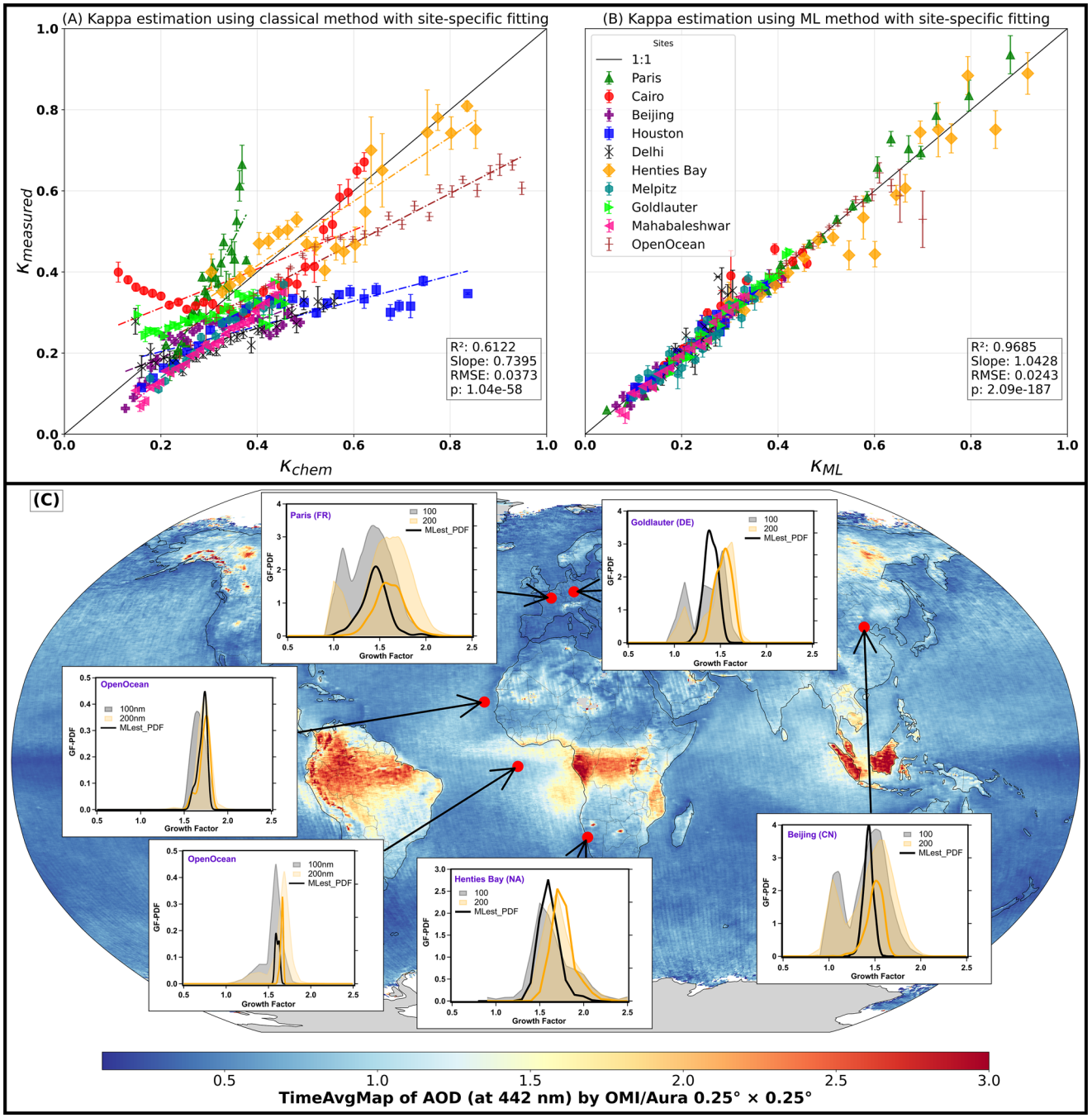

Fig. 2 | Comparison of hygroscopicity (κ) estimation and the influence of mixing state. A, **B** The κ estimation using a conventional bulk chemistry approach (κ_{chem}) and ML approach (κ_{ML}) with extreme gradient boosting (XGBoost) for all sites (ten diverse measurements) with a bivariate regression fit of binned data based on the geometric mean of κ bins with best-fitted diameters (all diameters (50–300 nm) are shown in Supplementary Figs. and tables). The threshold indicates the 95th and 5th data quartiles for the outliers in the geometric mean and standard deviations, with significance p-values of 1.04×10^{-58} and 2.09×10^{-187} for conventional and ML approaches, respectively. The interpretation of mixing state in (**C**) using GF-PDFs from measured and ML estimated hygroscopicity probability density (ML_{PDF}) function at six different unique sites (including open-ocean cruise measurements), with the availability of observed GF-PDF at 100 nm and 200 nm, considered as representative of smaller and larger particles, with the background of reanalysed global time average aerosol optical depth (AOD) from OMI/Aura GIOVANNI package[76]. These six sites are externally mixed dominant, which can largely influence hygroscopicity and grow particles bigger to $PM_{2.5}$, which may indirectly impact the parameters that drive AOD in the region[29], and are well represented by ML_{PDF}, also in a homogeneous/remote environment like Goldlauter (DE).

unresolved organic species, are not fully captured in chemical measurements. However, incorporating size distribution with aerosol chemistry, black carbon, and regional variability (ML framework shown in Fig. S5), the κ_{ML} captures these cases more robustly and may help resolve discrepancies caused by incomplete chemical speciation. The hygroscopicity parameter ($\kappa_{measured}$) computed from the measured size-segregated hygroscopic growth factor (GF) closely resembles κ_{ML} for all geographically diverse sites (Fig. 2B). Among ML models, extreme gradient boosting (XGBoost) provided the most consistent performance statistics, and it is based on random train–test partitioning (see methods and Fig. S9) and therefore reflects within-dataset predictive skill rather than temporally extrapolative performance.

Notably, site-specific behaviour further highlights these differences, such as Paris (suburban site, κ_{ML} with $R^2 = 0.92$; Fig. S2), which shows

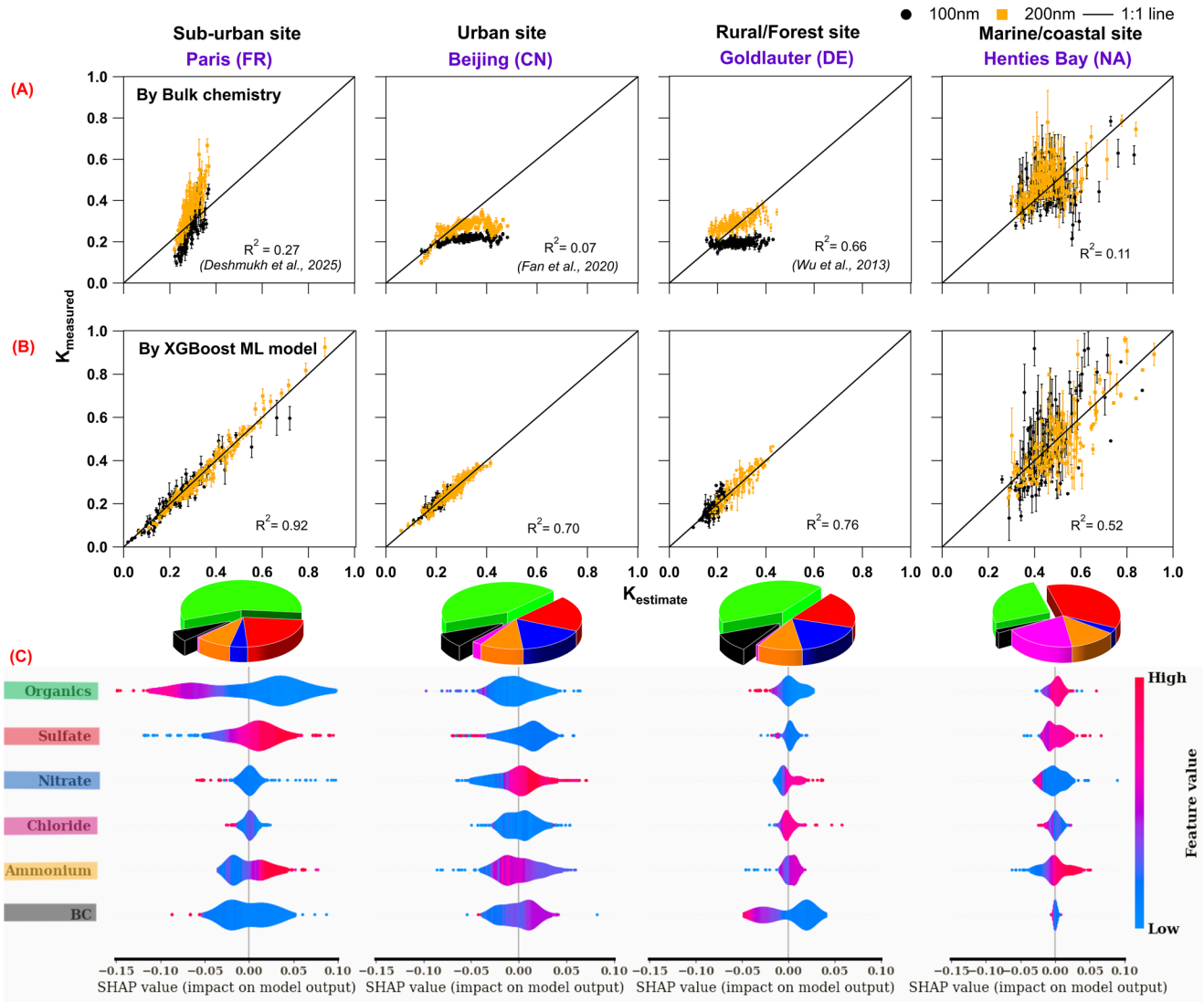

Fig. 3 | Comparison of hygroscopicity estimation and SHAP analysis. A, B The estimates of κ_{chem} and κ_{ML} (by XGBoost) are shown alongside $\kappa_{measured}$ in a scatterplot. The summary plots for the SHAP values are computed for test sets from different locations (**C**), with a pie chart showing the contributions of chemical species (i.e., Organics, Sulfate, Nitrate, Chloride, Ammonium, and BC). SHAP provides reliable insights when applied to models with strong correlations ($R^2 > 0.70$). However, at the Namibia site, where correlation is lower, interpreting potential driving-factor feature values should be approached with caution, though they still offer valuable trends. Detailed SHAP values with other variables (i.e., RH, ws, PNSD, etc.) are shown in Fig. S4 of the supplementary materials.

aerosol hygroscopicity primarily driven by sulfate and organic aerosols (OAs), which may be attributed to the increased oxygen content in the organic fraction (oxidised OAs), which facilitates water uptake[22,23]. While Henties Bay in Namibia (marine site, κ_{ML} with $R^2 = 0.52$; Fig. S2) aligns with high hygroscopicity for inorganic aerosols and chlorides transported over the site shown in Fig. S6B. Such marine aerosols are typically characterized by a mixture of secondary sulfate and organic aerosols, both quantified by aerosol mass spectrometry (AMS), which accounts for high hygroscopic growth[24]. In contrast, highly polluted environments (e.g., Delhi) display suppressed κ_{ML} correlation with measurements ($R^2 = 0.48$; Fig. S2), likely due to the dominance of low-hygroscopic components such as black carbon (BC) and primary organics except for specific events of high chloride[25] with complex mixing states and dynamic emission sources (Fig. S11). Lower estimation accuracy ($R^2 < 0.52$) at Henties Bay and Delhi is due to the fewer measurement data points at these locations, resulting in a smaller training dataset (Table S1), which further complicates model output. Furthermore, the Henties Bay site is influenced by marine aerosols, which remain challenging to fully quantify using AMS[26] and HTDMA[27].

Overall, this ML framework demonstrates applicability across diverse regional aerosol regimes, motivating further exploration of SHAP-based model interpretability analysis, which quantifies the relative influence of chemically and physically meaningful drivers on κ estimates and provides a physically consistent, model-based understanding of aerosol-water interactions.

Drivers of aerosol hygroscopicity

Incorporating aerosol mixing state into hygroscopicity estimates was possible using information from the aerosol size distribution. Aerosol size distribution reflects emission sources[27], capturing particle-mixing information beyond bulk chemical composition typically used in conventional models, thereby improving the representation of aerosol–water interactions. However, because size distributions also vary strongly across environments and may differ in representativeness and comparability between sites, their predictive benefit is not universal. Feature-combination tests show that PNSD inclusion improves predictive skill in specific regimes (e.g., Paris; $R^2 \approx 0.70–0.90$; Table S3), but not consistently across all sites. The improved performance of κ_{ML} reflects its capacity to capture uncharacterized chemical interactions between organic and inorganic fractions (Fig. S11), the role of BC density[28], uncertainties regarding the hygroscopicity of organics (κ_{org}), and the complexities of internal and external aerosol mixing states. SHAP analysis (Fig. 3C) and Growth Factor Probability Density Functions (GF-

PDFs) reveal that the inorganic-to-organic ratio and particle size, to some extent, shape κ across diverse environments. To evaluate these factors, we combined measured GF-PDFs[4,8,27] with ML-estimated distributions (ML_{PDF}; Fig. 2C) across six available measurements spanning suburban, urban, rural, open ocean, and marine environments.

In climate models, the ZSR mixing rule oversimplifies the multimodal growth behaviour commonly observed in ambient aerosols. Since κ_{chem} bulk composition lacks particle-level mixing information[11] and provides mean growth values exclusively[8], it is unable to capture the heterogeneity associated with externally mixed aerosol populations. These mixing states substantially influence κ and may indirectly affect aerosol optical depth (AOD) in the region[29]. As Gliß et al.[30] found, global models consistently underestimate total AOD by ~21% due to overestimating the scattering enhancement from κ[30] and models often assume an internal mixture, which erroneously assigns high water uptake to all particles. Similarly, the chemistry-derived distribution ($Chem_{PDF}$) and the measured GF-PDF exhibit apparent discrepancies (Fig. S6A). In marine-dominated environments such as Henties Bay, coarse-mode and externally mixed aerosol introduce significant variability in κ (0.1–1), where κ_{chem} correlates weakly with measured ($R^2 = 0.11$) compared to κ_{ML} ($R^2 = 0.52$; Fig. 3A). The chemistry-based (Fig. S7) approach performs better in homogeneous rural (e.g., Goldlauter) sites but systematically underestimates hygroscopicity in complex, mixed-source regions (Fig. 3A, B). While rural/remote sites may be well-mixed, 75% of global grid cells have seasonally averaged hygroscopic mixing state (χ) values between 20% and 63%, indicating external mixing[31].

Beyond κ estimation, SHAP identifies feature contributions within the trained model, provides insights into the potential drivers of ML estimates, and assesses whether these contributions are physically consistent across environments. Aerosol chemical composition emerged as the dominant driver of κ variability, followed by particle size distribution across sites (Fig. 3C and S4). In the SHAP framework, the SHAP value (x-axis) indicates whether a given variable contributes to increasing or decreasing κ for the given estimate, while the colour bar depicts the magnitude of the variables. The spread along the x-axis denotes each factor's contribution and importance; the greater the spread, the more important it is. For instance, the SHAP plot for Paris shows sulfate as the strongest positive contribution to κ (SHAP ≈ +0.10), consistent with the high hygroscopicity of sulfate-rich aerosols, while organic aerosol fraction and RH exert negative contributions (SHAP ≈ −0.15; Fig. S4), reflecting the suppressing effect of less hygroscopic organic components. In Beijing, nitrate emerges as a potential primary contributor to κ, underscoring the regional dependence and secondary inorganic aerosol formation in polluted environments[28]. Meteorological parameters and PNSD also influence the model and contribute to κ variability; although their effects vary across sites and therefore cannot be ignored, this highlights the site-specific aerosol interactions.

Overall, the SHAP analysis demonstrates that the ML framework attributes κ variability to physically plausible drivers and adapts to regional differences in aerosol composition and environmental conditions. Importantly, SHAP provides model-based attribution rather than causal inference, and evaluates the physical consistency and interpretability of the ML predictions[21,32]. Integrating κ_{ML} with SHAP-driven feature interpretation offers interpretable predictive accuracy. Unlike bulk-chemistry approaches that do not capture mixing states, ML techniques can adaptively account for spatial and compositional heterogeneity in aerosol populations. To extend the utility of κ estimation beyond individual sites, we also/then explore quantification of mixing state and diagnostic transferability analysis across diverse regions, identifying high-density or core points with structured subspaces.

Mixing state variability, improvements by ML, and limitations

The effect of mixing state (i.e., external mixing in near-source regions) on hygroscopicity depends strongly on the region's composition. However, basic relationships between chemical and physical factors remain largely consistent across environments and can be used in ML models. To assess, we quantify the mixing parameter across geographically and chemically diverse environments, using (see methods) a hygroscopicity-based proxy derived from growth-factor variability (χ_{GF}; Fig. 4A–F) and an independent bulk-chemical proxy derived from ACSM/AMS measurements (χ_{chem}; Fig. 3G, H). The pointwise error metric Δκ ($[\kappa_{chem}\text{-}\kappa_{measured}] - [\kappa_{ML}\text{-}\kappa_{measured}]$) indicates that ML predictions reduce absolute hygroscopicity errors across a wide range of mixing states, suggesting ML improvement over the κ_{chem} approach. The largest improvements occurring in externally mixed regimes ($\chi < 0.6$), which account for ~70–90% of particles across sites (Fig. 4). For example, Open Ocean, 38% of particles are towards intermediate to internal mixed regimes ($\chi > 0.6$), with ML improvements reaching $\Delta\kappa \approx 0.1$–0.3 (Fig. 4A), indicating that ML does not degrade predictive skill over internal mixing. At the Paris (suburban) site, ML reduces κ estimation errors across both externally and internally mixed regimes, with median improvements reaching $\Delta\kappa \approx 0.1$–0.2, the same as Houston (Fig. 4B, F). In contrast, at the Melpitz site, ML improvement shows higher variability (Fig. 4H), likely due to limited data availability and the relatively good performance of κ_{chem} for internal mixing. Despite these regional differences, ranging from open-ocean and remote continental locations to urban and suburban regions, ML performs well across externally mixed regimes and comparably robust in internally mixed regimes (Figs. 4, S12G, H). This aligns with previous studies[31], showing that a substantial fraction of aerosol populations (70–80%) resides in partially externally mixed regimes ($\chi < 0.6$; Fig. 4). Overall, this indicates that the ML framework adapts to regionally varying mixing states.

This regime-dependent behaviour motivates the question of whether relationships learned in one chemically stable environment can be applied, at least in part, to other environments with similar or non-similar aerosol characteristics. To evaluate this possibility, we conducted a diagnostic transferability analysis in which an ML model trained in one location (the source site) was applied to an independent location (the target site) that was entirely excluded from training. Note that these experiments assess only the sensitivity of κ estimates to cross-site application under constrained conditions, rather than validating the generalisation and transferability of the entire ML framework. Due to data limitations and regional variability across sites, achieving full transferability is difficult and is out of scope for this study.

To test the potential cross-site applicability and to simplify the model as much as possible, while ensuring consistent input features, we use only the "chemical variables" as model inputs, and all variables are standardized consistently across source and target locations. This simplification intentionally constrains the analysis to composition-driven methods and also reduces bias from site-specific variables (e.g., meteorology or PNSD) that are not uniformly available across sites. Although we train the model with PNSD for potential transferability, this further complicates the model and does not consistently improve performance across sites (see Fig. S12A–F). This indicates that PNSD provides conditional added value, particularly for within-site interpretation, while its utility for cross-site transfer is limited by site-specific variability. We further evaluate a single model trained on combined data from two or more sites, reproducing κ variability but with comparable lower predictive skill (Melpitz; $R^2 \approx 0.56$; Table S5). Further, three representative sites were used for transfer cases that span distinct environments (Fig. S13). When a model trained on continental source sites from Houston, Henties Bay, Goldlauter, and Melpitz to the Open Ocean (R/V Polarstern) yields only moderate ($R^2 \approx 0.2$) overall agreement with $\kappa_{measured}$. However, a dense subset of data points shows substantially better agreement ($R^2 \approx 0.70$; Fig. S13A), indicating that the ML model captures κ reliably within chemically stable marine regimes. Larger deviations occur outside this core region, suggesting that episodic influences or compositionally distinct aerosol populations are not well represented by the source-site training data, highlighting that cross-site applicability over the open ocean is regime-dependent. Similar behaviour is observed when transferring a model to other sites, which shows structural consistency rather than individual-point accuracy (Supplementary Note 1 and Fig. S13).

There are still challenging issues that could be addressed to provide ML-based large-scale estimates of aerosol hygroscopicity. The possibility of

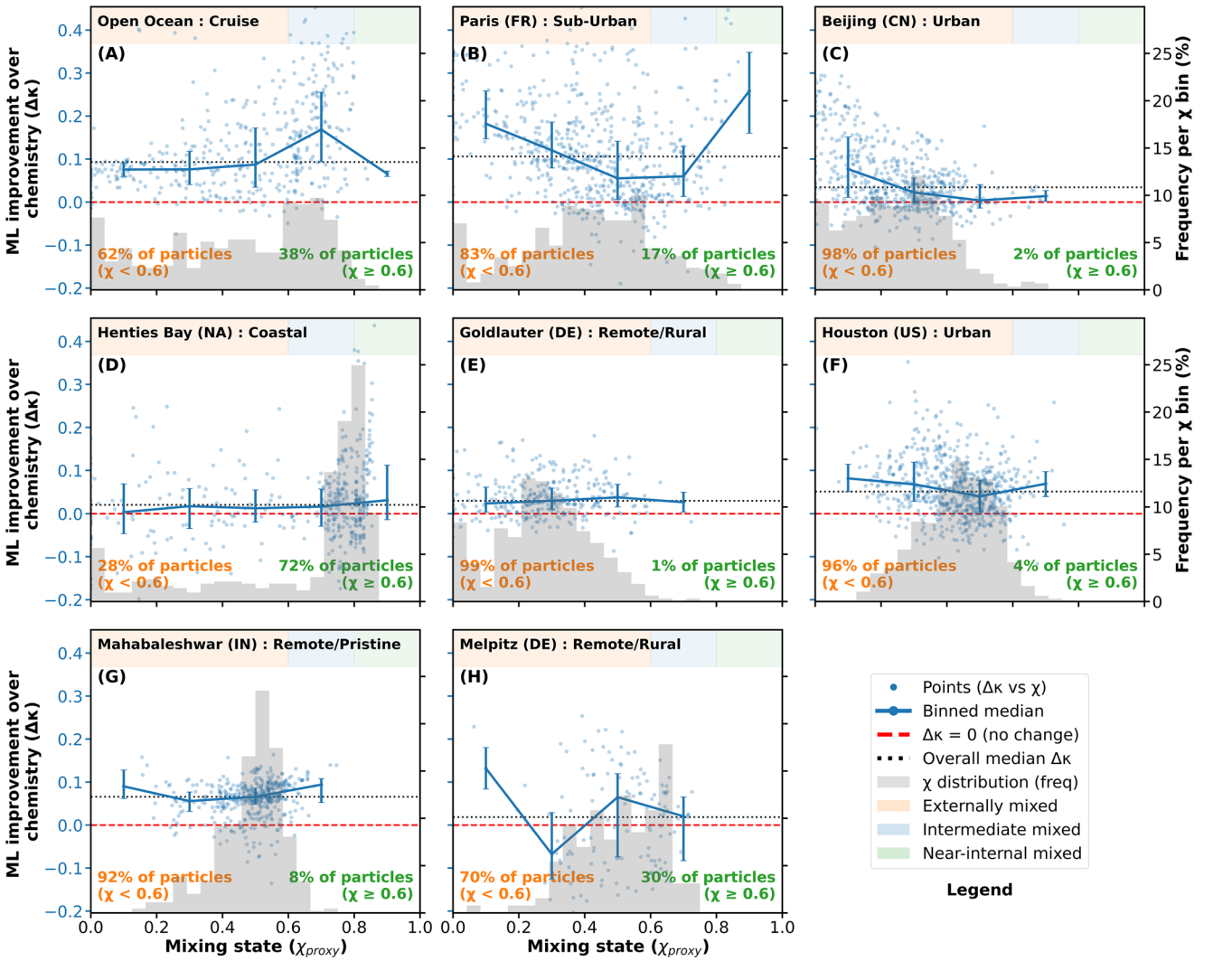

Fig. 4 | Machine-learning based hygroscopicity performance as a function of aerosol mixing state. Dependence of κ_{ML} prediction performance on aerosol mixing state across diverse environments. **A**–**E** show a hygroscopicity-based mixing-state proxy χ_{GF} derived from HTDMA growth-factor spread for OpenOcean (Polastern cruise), Sub-Urban, Urban, Coastal, and Pristine remote environments. **F**–**H** show a bulk chemical proxy χ_{chem} derived from chemical composition-based chemical diversity Dγ[76] for Houston and remote sites like Mahabaleshwar (IN) and Melpitz (DE). The panels show the pointwise reduction in absolute hygroscopicity error, $\Delta|\kappa|$, relative to bulk-chemistry closure, with grey histograms indicating the frequency of mixing-state conditions. The corresponding gain in explained variance, binned by mixing state (χ_{GF}) or chemical diversity (χ_{chem}) on the x-axis as χ_{proxy}. Across all sites and both independent proxies, ML provides the largest benefit under chemically complex or externally mixed conditions, while exhibiting comparably good performance to bulk-chemistry closures in near-internal-mixing regimes.

transferring a model trained at one site (where HTDMA instrumentation is available) to a similar or dissimilar site without HTDMA instrumentation to retrain the model is difficult to achieve at this point. This scenario concerns the potential transferability of the models and constitutes a research field in itself, where domain adaptation and concept drift mitigation techniques are promising for improving model generalization. Approaches such as meta-learning and few-shot learning enhance transferability by enabling models to adapt to new sites using only limited local data, thereby facilitating future studies. Here, we identify high-density points as structured subspaces with higher transferability potential (Supplementary Note 1, Fig. S13). It should also be noted that, although the presented data sets are novel and have wide geographical coverage, the limited number of samples, the limited time span, and the heterogeneity make it difficult to train transferable models. Nevertheless, this is a very relevant line of research in aerosol science and artificial intelligence, seeking solutions to overcome the limitations mentioned above and to provide a general ML framework that can be used worldwide. The main evaluation framework adopted here is most appropriate for assessing κ estimation and gap-filling within the represented data domain. The sequential validation conducted for Houston (R^2 = 0.75 (random split) to R^2 = 0.60 (time-blocked split); Fig. S3) shows that performance decreases under a more rigorous temporally structured setup. Thus, the present framework should be interpreted as a regionally informed and interpretable modelling approach, rather than a universally transferable predictor across all atmospheric regimes.

Overall, this analysis demonstrates that transfer of ML-based hygroscopicity estimates across sites is feasible only within limited, compositionally similar regimes, although generalization transferability remains challenging. Differences in data distributions[33], sampling periods, and aerosol sources strongly influence transfer performance, as reflected in site-to-site variability[21] (see Figs. S10–S11 for heterogeneity in chemistry). These results highlight both the potential and the limitations of model transferability and underscore the need for regionally informed training data. In this study, we use a single input for testing transferability, namely the chemical composition, although it exhibits its own variability and non-collinearity[33]. Further improvements, tuning, and site-specific limitations must be considered when evaluating model transferability, including outlier removal, parameter selection based on collinearity, and gap-filling for values below the κ-based threshold. Understanding these can be essential for further steps

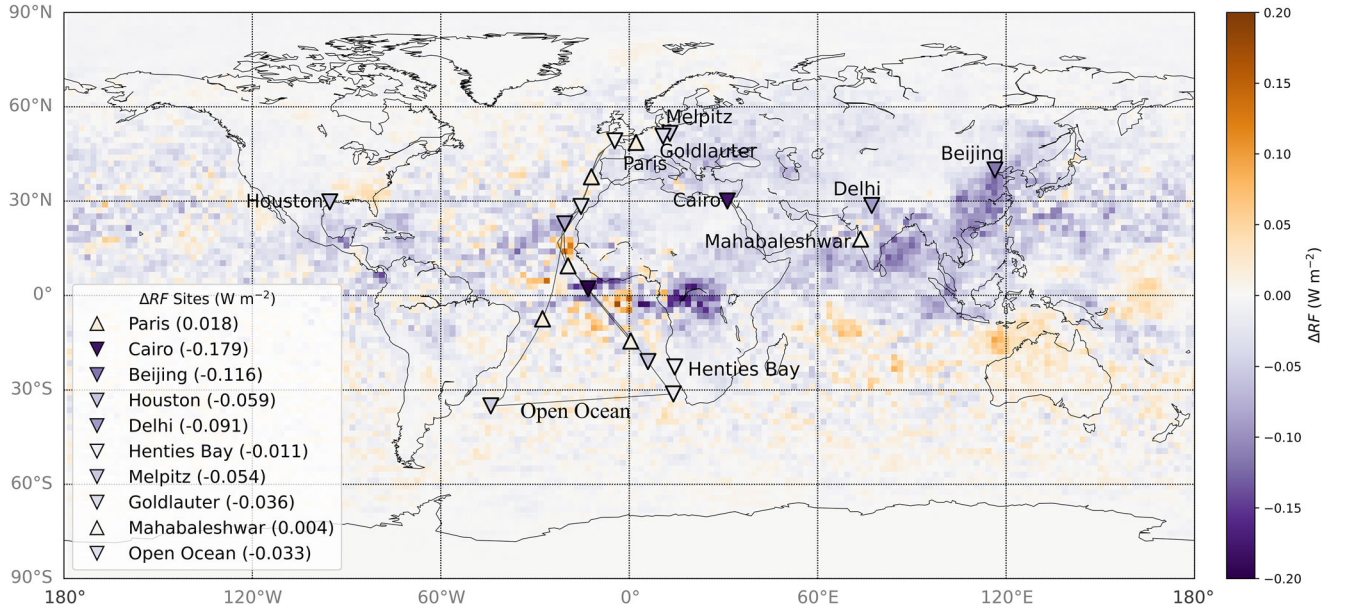

Fig. 5 | First-order sensitivity assessment on radiative forcing calculation when implementing κ_{ML} at the measurement sites. Shown is the relative change in top-of-atmosphere shortwave aerosol–radiation interaction (ΔRF_{ari}) estimated by replacing the default hygroscopicity (κ) used in the ECHAM–HAM model with ML-derived κ at the measurement sites, indicated by ΔRF mean values (listed in the legend). These site-level estimates are calculated using a pre-computed radiative forcing response to hygroscopicity (ξ; see Methods and Fig. S8) and represent a first-order, analytic quantification of how regionally varying κ may influence radiative forcing. Cruise track is shown in a grey dashed line for Open Ocean (R/V Polarstern), and ΔRF is evenly spaced (whole cruise; see Fig. S8A). The background map is based on the ΔRF results, adapted from Pöhlker et al.[12] and is included for contextual comparison only of the site-specific ML results.

in revising estimations of hygroscopicity parameterizations in climate models, particularly for externally mixed aerosols and constraining uncertainties in aerosol-radiation feedback.

Change in direct radiative forcing using ML estimates of hygroscopicity

Analytical incorporation of improved κ predictions into radiative forcing calculations reveals a stronger aerosol cooling effect, suggesting a potentially substantial impact on direct radiative forcing (RF_{direct}). The change in radiative forcing (ΔRF) was obtained using a pre-calculated sensitivity to aerosol hygroscopicity (ξ), based on simulations from the aerosol-climate model ECHAM–HAM, which is the ECHAM atmospheric general-circulation model coupled to the HAM aerosol module. It is acknowledged that directly replacing aerosol schemes in climate models and running simulations with ML outputs is computationally complex and technically challenging; hence, using this simplified ΔRF calculations approach (see Methods) based on ξ (Fig. S8B) provides a first-order estimate of the impact on RF (shown in Fig. 5). The ΔRF in the shortwave range aerosol-radiation interaction (ΔRF_{ari}) at the top of the atmosphere (TOA) is determined using ξ and Δκ, which represents the difference between the default model reference κ and the improved κ_{ML} for our study locations (Eq. 12). There are limitations to this simplified ΔRF approach, which assumes a linear relationship between κ and ΔRF; however, it was estimated using the empirical framework of Liu et al.[34], which shows that ΔRF depends approximately linearly on κ ($R^2 \approx 0.81$), even at locations exhibiting large Δκ (Supplementary Note 4 and Fig. S8C and S14). Over the observed κ range, ΔRF increased approximately linearly with κ within individual sites, whereas all sites introduced additional scatter due to differing baseline forcing conditions. This study demonstrates the potential influence of machine-learning-based κ parameterization on RF estimates and does not provide absolute values of net radiative forcing.

Use of κ_{ML} values indicates a negative impact on ΔRF (-0.059 ± 0.062 W m^{-2}), showing a larger cooling effect than the case where the conventional method of κ prediction was used[12] with no external mixing accounted for (-0.026 ± 0.022 W m^{-2}), indicating that the ML-driven hygroscopicity estimation captures the change in RF (Table S4). Interestingly, the estimated ΔRF at Paris shows a positive (0.018 W m^{-2}) trend in Fig. 5, which is mathematically consistent with Eq. 12. At Paris, both the sensitivity (ξ is −0.169; Fig. S8) and Δκ (−0.104) are negative, resulting in a positive ΔRF. This could imply that replacing the conventional κ with κ_{ML} reduces aerosol water uptake and light scattering at Paris, thereby weakening the cooling effect. While site-specific uncertainties, such as those in Henties Bay, may introduce localized biases, their influence on the global forcing pattern is limited. These differences are significant by a factor of 2 across urban, suburban, rural, coastal, and open-ocean regions when models localised κ (Table S4). Ignoring this information, such as mixing state and chemical variability, could lead to inaccuracies in estimating RF and its climatic implications.

The findings of this study reinforce the need to incorporate site-specific and regional aerosol data (such as κ_{ML} values and their impact on RF) into global climate models for better accuracy. Global models often rely on a general chemistry module that cannot resolve mixing state, but integrating more localised insights could lead to improved predictions of how aerosols contribute to cooling or warming. This emphasises the importance of refining global models by incorporating sensitivity analyses and heterogeneous aerosol properties from diverse environments. Incorporating complex ML equations in global models is technically challenging. Still, it can be the first step for the next-generation climate models, moving from regional to global scales. Further studies can explore how localised hygroscopicity estimates (such as those from our sites) contribute to regional climate phenomena (e.g., monsoons, droughts) and global climate trends.

Summary, implications, and significance

Aerosol hygroscopicity (κ) in the sub-saturated regime plays a key role in regulating particle growth, light scattering, and water uptake, thereby influencing the atmospheric radiative budget. Complex aerosol mixing is typically found in regions with large local pollution sources, where a significant number of people reside. Better knowledge of κ will not only improve estimates of climate effects, but also fog formation and health effects[35]. Yet, a limited understanding of regional variability in κ continues to hinder accurate representation of aerosol mixing states and radiative forcing in climate models. To address these challenges, an ML framework was

developed to provide a computationally efficient approach that enables accurate imputation of missing κ values[20], providing high-resolution, size-resolved, geographically diverse region-specific κ estimates. Although a generalizing model transferability lies beyond the present scope, future progress will benefit from expanding long-term global aerosol datasets and ongoing advances in machine learning and artificial intelligence (AI)[21]. These site-specific models capture local relationships among composition, mixing state, and hygroscopicity, while cross-site models test these relationships across different environments. Limitations arising from physically distinct aerosol regimes, reduced performance in cross-site applications, and not model overfitting. This work is limited to investigating how region-specific aerosol properties influence κ prediction and model performance, not the global predictor framework.

Building on previous work that emphasised the importance of representing external mixing in climate models[12], our regional κ_{ML} estimates provide an improved, data-driven depiction of aerosol hygroscopicity. This approach indicates more pronounced negative radiative forcing estimates and offers an alternative to conventional uniform parameterisations that assume fixed hygroscopicity parameter values based on individual aerosol components[29]. Chemically complex and polluted regions, such as Delhi and Cairo, likely experience stronger hygroscopic growth and water uptake[36,37], which may partly account for the observed regional cooling trends or slower warming over the Asian and African continents[38] respectively. Enhanced hygroscopic growth in such regions also carries potential implications for public health, though this remains beyond the present scope[35]. These findings highlight the importance of data-driven region-specific aerosol parameterisations as a critical step toward reducing uncertainties in estimates of direct radiative forcing in next-generation climate models. The use of κ_{ML} estimates can perturb regional RF_{direct} by typically up to ±0.1 W m^{-2} with a general shift toward stronger cooling on most of the selected measurement sites, which may become significant on a global scale when the κ_{ML} algorithm is integrated into global models, potentially altering both the magnitude and the sign of aerosol–radiation interactions.

This framework provides a scalable tool for improving estimates of aerosol–radiation interactions and guiding the design of targeted observations. Our results reinforce the need to incorporate regional aerosol variability into climate models and highlight the value of expanding observational datasets in the sub-saturated regime. Emerging ML and AI approaches offer powerful means to address data gaps and uncover key patterns in aerosol behaviour globally, marking a step toward more accurate and physically grounded representations of κ in future climate assessments.

Methods

Size-resolved HTDMA data and aerosol mass spectrometry

The aerosol mass spectrometer (AMS) and size-resolved humidified tandem differential mobility analyser (HTDMA) data used in this study were obtained from ten field studies summarised in the next section. AMS data on the non-refractory submicron aerosol fraction (i.e., mass concentrations of organics, nitrate, sulfate, ammonium, and chloride) were obtained either from an aerosol chemical speciation monitor (ACSM, Aerodyne Research Inc., Billerica, MA, USA) or a high-resolution time-of-flight aerosol mass spectrometer (HR-ToF-AMS, Aerodyne Research Inc., Billerica, MA, USA). Note that the AMS measurement of Cl^- from sea salt is typically not quantitative due to the partial evaporation of the refractory NaCl as a primary compound of sea salt aerosols[39]. The concentration of equivalent black carbon (BC) was measured using a multi-wavelength aethalometer[40] (AE33 model, Magee Scientific, 1-minute time resolution, wavelength λ = 880 nm). For ambient measurements, uncertainties in BC concentrations using the AE33 can range from 5% to 15%[41].

An HTDMA was used to measure the ambient aerosol hygroscopic growth factor and distributions[42]. It is to be noted that the variance of RH may have a measurable impact on GF, particularly with low-soluble compounds and their sensitivity to RH in this range. This could contribute to GF values uncertainty, especially for organics or less-hygroscopic dominant species[8]. Although the RH calibration was maintained carefully, this uncertainty is acknowledged as a potential source of error in our measurements from different sites of HTDMA data and presented in previous studies through corresponding original publications[14,25,28,37,43–48]. The HTDMA sizes used here (50-300 nm) partially capture the scattering contribution from submicron aerosol but do not capture the contribution from the coarse mode. Thus, light-scattering estimates based solely on HTDMA data generally underestimate total light scattering[49]. The particle number size distributions of ambient aerosol in its dry state were measured using a Scanning Mobility Particle Sizer (SMPS), operated concurrently with the HTDMA and AMS at most stations (see Table S1).

Field studies and data sets

This study integrates nine field campaigns conducted worldwide (see Table S1), during which size-resolved HTDMA, AMS/ACSM, BC, and SMPS measurements were performed.

i) **Beijing, China:** Two field campaigns were conducted (Fig. S10) during the winter of 2016 and summer of 2017 in urban Beijing (39.97° N, 116.37° E; 50 m a.s.l.) to measure aerosol physical and chemical properties (Fig. S10 & S11)[28]. The BJ site is located at the Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences, situated between the North Third Ring Road and the Fourth Ring Road in northern Beijing. The sampling period in the cold season was from 16th November to 10th December 2016, during the domestic heating period in Beijing. The sampling period in the warm season was from 25th May to 18th June 2017, with an average of 1300 data points[28].

ii) **Cairo, Egypt:** Atmospheric measurements in the frame of the POLCAIR project were performed in Cairo, the capital city of the Arab Republic of Egypt, located at the base of the Nile Delta. A comprehensive suite of state-of-the-art aerosol instrumentation was deployed at the premises of the National Research Centre of Egypt (30.04° N, 31.23° E; 250 m a.s.l.). The sampling site is located approximately 3.5 km southwest of the city centre. The field campaign started on 27th November 2019 and ended on 29th January 2020, resulting in more than two consecutive months of continuous observations[37] of HTMDA, SMPS, and ACSM measurements. This study provided an average of 1000 data points.

iii) **Delhi, India:** Real-time measurements of atmospheric aerosols were conducted[25] during winter (1st February to 20th March 2020) at Block 5 of the Indian Institute of Technology Delhi (28.61° N, 77.20° E; 200 m a.s.l.). HTMDA, ACSM, and Particle number size distributions (PNSD via SMPS) were measured, resulting in ≈400 data points.

iv) **Goldlauter, Germany:** Physical and chemical characterization of the atmospheric aerosol was performed in the mid-level mountain area in Thuringia, central Germany, nearby the village of Goldlauter (50.38° N, 10.45° E, 605 m a.s.l.) from September to October 2010 as part of the Hill Cap Cloud Thuringia 2010 (HCCT-2010) experiment with ≈1400 data points for ML framework in this study. Hygroscopicity, chemical composition, and PNSD were measured using HTMDA, AMS, and SMPS[44].

v) **Henties Bay, Namibia:** Physical and chemical characterization of the atmospheric aerosol was performed at a coastal site of the University of Namibia in Henties Bay, Namibia (22.63° S, 14.46° E; 20 m a.s.l.) in the framework of the AErosols, RadiatiOn and CLOuds in southern Africa (AEROCLO-sA) project[46]. Instruments were located in the PEGASUS mobile facility operated at the site between 26th August 2017 and September 10th, 2017. Particle hygroscopicity was measured with a custom-built HTDMA[45]. Aerosol non-refractory composition in the accumulation mode was measured with a QTof-AMS[50]. This study provides ≈920 data points.

vi) **Houston, USA:** During the TRACER campaign, a trailer was stationed at the main site of the Department of Energy's Atmospheric Radiation Measurement (ARM) program's ARM Mobile Facility (AMF1) in La Porte, Texas (29.76° N, 95.36° W; 10 m a.s.l.). The campaign was scheduled for one year, from September 2021 to September 2022[14]. An intensive observational period occurred between 29th June and 29th

Table 1 | Gravimetric densities ρ and hygroscopicity parameters κ at the sub-saturated regime used in this study[5,42,43,55,69–75]

Species	NH_4NO_3	H_2SO_4	NH_4HSO_4	$(NH_4)_2SO_4$	NH_4Cl	NaCl*	Organic matter	Black carbon
ρ (kg m^{-3})	1720[(a)]	1830[(a)]	1780[(a)]	1769[(a)]	1520[(d)]	2160[(e)]	1400[(a, c)]	1770[(a, b)]
κ	0.58[(a)]	0.9[(a, b)]	0.56[(a)]	0.48[(a)]	1.02[(d)]	1.1[(e)]	0.1[(a, c)]	0

a(5,68); b(68,69,70); c(70,71); d(72,73); e(74,75); *only used for Henties bay and Open Ocean (R/V Polarstern) due to data availability.

September 2022; ≈1100 data points. The chemical composition was measured using an AMS/ACSM, and HTDMA was used to determine hygroscopicity at four diameters (100, 150, 200, and 250 nm). The PNSD measurements using SMPS data were available for a limited period and had missing data gaps (which were also accounted for in the estimations). This may further affect the correlations in the ML framework.

vii) **Mahabaleshwar, India:** High Altitude Cloud Physics Laboratory (HACPL), Mahabaleshwar, over the Western Ghats region in the Maharashtra state, India (17.92° N, 73.66° E, 1378 m a.s.l.), established in the year 2012 with a basic suite of instruments such as aerosol size spectrometer, CCN counter, etc. The site is one of the highest rainfall receiving stations in the country during the summer monsoon season due to the orographic influence of strong moisture-rich south-westerly winds from the Arabian Sea[43]. Different air mass flow patterns influence the site in different seasons. The measurement from 01st February 2020 to 15th March 2020, ≈1400 data points. Hygroscopicity (for diameters, see Table S2), chemical composition, and PNSD were measured using HTMDA, AMS, and SMPS, respectively.

viii) **Melpitz, Germany:** The atmospheric aerosol measurements were carried out at the TROPOS research station at Melpitz (51.54° N, 12.93° E; 86 m a.s.l.), located approximately 50 km northeast of Leipzig, Germany[47,51]. The station is mainly encircled by agronomical pastures and forests within a rural area, so it is recognized as a rural background station[52]. The field campaign, as MELCOL, started on 20th May 2015 and ran until 02nd July 2015, resulting in ≈500 data points. Hygroscopicity, chemical composition, and PNSD were measured using HTMDA, AMS, and SMPS.

ix) **Paris, France:** The site is located in a suburban area approximately 23 km southwest of the Paris city center (2.148° E, 48.708° N; 150 m a.s.l.), and the measurement was conducted during the ACROSS campaign (Atmospheric ChemistRy Of the Suburban Forest, in Paris, Summer 2022) from 15th June 2022 to 17th July 2022[8], collecting ≈1800 data points. The particle's hygroscopic growth at 90 % relative humidity (RH) and chemical composition were measured using an HTDMA (scanning at 100, 150, 200, and 250 nm) and HR-Tof-AMS with SMPS for size distribution.

x) **Open Ocean:** Online measurements using a series of instruments (HTDMA at 100, 200, 300 nm, HR-Tof-AMS, and SMPS) were performed on the R/V Polarstern in three Atlantic cruises in spring and autumn of 2011 and 2012, collecting ≈1500 data points. These open-ocean voyages were between Bremerhaven, Germany, and Cape Town, South Africa, or Punta Arenas, Chile. For the three cruises (CR), CR1 and CR2 had almost identical ship tracks, with the starting harbour swapped with a ship track similar to those of CR1 and CR2 in the northern part (15−55°N). More details of the same Polarstern cruises can be found in a previous paper[53].

Details of the experimental conditions and measurement procedures for these campaigns are provided in the original studies.

Retrieval of hygroscopicity parameter κ, bulk chemical-derived κ_{chem}, and GF-PDFs

Retrieval of hygroscopicity parameter κ. Hygroscopic properties of aerosols can be described by the growth factor (GF), which is most commonly measured by a hygroscopicity/humidified tandem differential mobility analyser (HTDMA) instrument in the sub-saturate regimes[5]. The GF determined by the HTDMA is the ratio of the particle mobility diameter, D_p (RH), at a given RH (90%) to its dry diameter, D_0, in Eq. (1):

$$GF(RH,\ D_0) = \frac{Dp\,(RH)}{D_0} \tag{1}$$

The hygroscopicity parameter (κ) can be computed using the hygroscopic growth factor (GF) measured through an HTDMA[5,54].

$$\kappa_{measured} = (GF^3 - 1)\left(\frac{\exp(\frac{a_\omega}{D_0 \cdot GF})}{RH} - 1\right) \tag{2}$$

where a_ω is the water activity at which the growth factor was measured. According to Köhler's theory[55], a_ω is obtained by the following:

$$a_\omega = \frac{4\,\sigma s/a\,Mw}{RT\rho w} \tag{3}$$

Where $\sigma_{s/a}$ is the droplet surface tension (assumed to be that of pure water, $\sigma_{s/a} = 0.0728$ N m^{-2}), M_w is the molecular weight of water, ρ_w is the density of liquid water, R is the universal gas constant, and T is the absolute temperature.

Alternatively, κ_{chem} can be predicted using a simple mixing rule based on chemical volume fractions ($V_f i$), as proposed by[6]:

$$\kappa_{chem} = \sum\nolimits_i V_f i.\,\kappa_i \tag{4}$$

Here, κ_i and $V_f i$ are the hygroscopicity parameter and volume fraction for the individual (dry) component in the mixture, with i the number of components in the mixture. The respective sites used the chemical composition from the AMS, ACSM, and AE33 (black carbon) measurements. The approach we used to predict κ_{chem} follows the Zdanovskii-Stokes-Robinson (ZSR) mixing rule. To some extent, we indirectly incorporate solubility and chemical composition information into κ calculations, particularly for particles with organic fractions exceeding 50%. Hence, in ZSR-based κ predictions, the values of κ_i used in this study (Table 1) are based on literature-measured values with some water activity calculation[8] and uncertainties in the datasets are already accounted for. 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%[56,57].

GF-PDF estimations. Where the κ_{chem} and ML estimated GF-PDF is calculated using the corrected GF by Sjogren et al.[27]

$$GF(a_\omega, \kappa) = \left(1 + k\frac{a_\omega}{1 - a\omega}\right)^{1/3} \tag{5}$$

First, we calculated the bulk κ_{chem} from the ZSR ion mixing rule, and then the corresponding corrected growth at 90% was calculated using Eq. 5. After converting κ_{chem} to GF_{chem}, the frequency distribution and the probability density function (PDF) were calculated as in ref. 8; this conversion is needed when comparing with the measured GF-PDF. The frequency distribution provides the number of GF_{chem} values within specified intervals (bins). Kernel density estimation (KDE) and a non-parametric method that smooths data by summing Gaussian functions (Silverman's rule of thumb) provide a continuous representation of the distribution,

which is commonly used to analyse aerosol GF distributions. A similar method was used to obtain the ML-estimated PDF (ML_{PDF}) shown in the results above (Fig. 1C).

Machine learning models and data processing

Machine learning techniques have recently been used to estimate aerosol hygroscopicity[20]. Supervised regression techniques have generally been leveraged to implement the so-called proxies[58,59]. An ML-based proxy is an ML model that estimates a phenomenon or parameter using indirect measurements from other sensors. In the context of this article, we used two complementary predictor configurations; regression techniques are used to estimate aerosol hygroscopicity (κ) from chemical composition (including BC) and particle number size distribution (PNSD) measurements along with local meteorological parameters to evaluate the maximum within-site predictive skill and to examine the contribution of different predictor classes to κ variability. Specifically, the κ estimate $\hat{y}_t$ given some features $\boldsymbol{x_t}$ at a given time step t is set to:

$$\hat{y}_t \approx f_\kappa(\boldsymbol{x}_t) \tag{6}$$

ML Function $f_\kappa(\cdot)$ is used to estimate κ from indirect measurements $\boldsymbol{x_t}\{\subseteq \boldsymbol{x}_{pnsd}, \boldsymbol{x}_{chem}, \boldsymbol{x}_{met}\}$. To obtain the function $f_\kappa(\cdot)$, ML models from different families (linear, kernel, and ensemble) have been employed in the ML pipeline (Fig. S5). As explained in the following subsections, each model has its own hyperparameters and needs to find those that produce a good model. In parallel, a reduced-feature configuration based on variables that were more consistently available across sites (e.g., chemical composition) was used for cross-site comparisons and to test potential transferability. This separation was adopted to ensure that transfer experiments were conducted using harmonized predictor sets, while still allowing assessment of the added value of more physically informative variables under site-specific conditions. Furthermore, transferability and generalization are outside the scope of this study due to data limitations and regional variability.

The various ML models are tested and executed to estimate κ. Extreme gradient boosting (XGBoost; Fig. S9) performs and predicts κ well, underscoring its robustness in resolving hygroscopic behaviour across different locations and compositional variability, compared to other models: Random Forest Regression (RFR), Multiple Linear Regression (MLR), and Support Vector Regression (SVR). When training models for κ estimation, there are two alternatives: creating a separate model for each site and diameter, or combining data from different sites to create a global model. In this article, we have opted for the first option, since, as seen in the SHAP interpretability analysis (Fig. S4), the data set is heterogeneous in terms of time period and different sites, with different influencing factors. Indeed, we conducted an experiment to compare the estimation performance of a site-specific model with that of a model combining data from two similar sites (Table S5). The results showed that site-specific models performed better.

The model output across distinct atmospheric conditions is depicted in Fig. S2, which shows the best-fit diameter correlation between estimated and measured hygroscopicity (HTDMA) at different sites, and other diameter analyses are briefly described in Tables S2 and S3. The framework was evaluated across the full range of available site-specific diameters (50–300 nm); detailed metrics for all sizes are provided in the Supplementary Tables 1–5 and Supplementary Figs. 1–5.

Data split and model selection. The data is randomly split into training and test sets at 70% and 30%, respectively. This division mechanism achieves good model performance and simulates the application of missing-value imputation. As an alternative, a sequential temporal split was tested on the dataset with the largest time span (the Houston, USA dataset), selecting the first 70% for training and the remaining 30% for testing. Figure S3 compares the performance of random and sequential splits over time, demonstrating that the random split performs well and is appropriate, as it allows the model to learn the full range of variability in the dataset and is consistent with previous studies (e.g., Vu et al.[20]), whereas a time split is more rigorous and requires extensive data sets. Random splitting is appropriate for assessing within-dataset interpolation performance, but it does not constitute a test of temporal extrapolation. Further, it is used to assess the model's ability to capture relationships within the available dataset. The training data is used to find the best hyperparameters and train the model. For each ML model, a grid search is performed on the training set using 10-fold cross-validation (CV) to optimize the hyperparameters Θ. In transferability experiments, the only difference lies in the data used for training and testing. In the transferability experiment, the model was trained on data from one site (or several mixed sites) and tested on data from another site, which was not included in the training set.

Data pre-processing. In the first stage of feature engineering, the variables to be used in the estimation are selected. Due to the PNSD variable's high dimensionality, principal component analysis (PCA) performs a dimensionality reduction, selecting the number of latent components that maintain at least 90–95% of the variance of the original data. In fact, by using PCA, the data is assumed to reside in a low-dimensional space spanned by certain bases and is represented as a linear combination of orthonormal vectors. The PNSD measurements are projected onto the subspace spanned by the set of leading principal components. Finally, all measurements are standardized to promote training stability. Further, using PCA might remove sites with different instruments for PNSD (e.g., different mobility ranges, APS vs. OPC in future studies), where the initial size distribution is on a fixed grid without disturbing the multi-lognormal distribution in processing. It should be noted that PCA is performed to avoid more complexity in the ML framework.

ML models. Four different ML models are evaluated for κ estimation. First is the MLR, which assumes a linear relationship between the input and target variables. Second is the RFR, an ensemble method whose prediction corresponds to the average prediction of a group of decision trees. In addition, the decision trees follow a decorrelation process to improve the ensemble accuracy. Third, the SVR model, the equivalent of the support vector machine for regression, implicitly uses kernels and the epsilon-insensitive loss to map the data into a higher-dimensional space. Fourth, the XGBoost, which is a tree-boosting method that combines the output of W weak learners (decision trees), resulting in the estimator[60]:

$$\hat{y}_t = f_\kappa(\boldsymbol{x}_t) = \sum_w f_\kappa^w(\boldsymbol{x}_t) \tag{7}$$

XGBoost uses a gradient boosting strategy, where each weak learner is trained sequentially to improve the error of the previous learner. A second-order Taylor expansion is used in the loss function to follow this iterative procedure effectively. Some XGBoost hyperparameters include the number of weak learners, the maximum decision tree depth, and the learning rate. After training the models with the best hyperparameters, the estimates $\hat{y}_\kappa$ can be obtained for the desired samples, e.g., in the case of missing values, estimations are obtained for the data gaps.

Interpretability and factor discovery. Lately, explainable artificial intelligence (XAI) techniques have enabled interpretable results from so-called black-box ML models. To address the challenges associated with the interpretation of complex, high-dimensional ML predictions, Shapley additive explanations (SHAP) are implemented. SHAP is an interpretable ML tool that provides insights into the relative influence of various input features on model predictions[32]. Recently, SHAP has been leveraged to determine input variables' contributions and potential driving factors in the ML-based estimation of environmental parameters[61,62]. Henceforth, the ML pipeline presented includes an explainability module that provides interpretability using SHAP values. SHAP provides an exact way to compute Shapley values, a concept

borrowed from game theory, where in a cooperative game, the Shapley values are the average marginal contribution of each of the ML model input variables across different coalitions[32]. The resulting explainability model can be seen as an approximation of the model f_κ by the additive model g_κ[32]:

$$f_\kappa(\mathbf{x_t}) = g_\kappa(\mathbf{x_t}) = \phi_o(f_\kappa) + \sum_{i=1}^{m} \phi_i(f_\kappa, \mathbf{x_t}) \quad (8)$$

Where $\phi_i(f_\kappa, x_t)$ is the SHAP value for the i-th feature, and $\phi_o(f_\kappa)$ is the average prediction ($f_\kappa(0)$). The SHAP values $\phi_i(f_\kappa, x_t)$ can be computed as[32]:

$$\phi_i(f_\kappa, x_t) = \sum_{s \in S} \frac{1}{P!}\left[f_\kappa\left(x_t^{P_i^s \cup i}\right) - f_\kappa\left(x_t^{P_i^s}\right)\right] \quad (9)$$

Where P is the number of input variables, S is the set of all feature orderings, $\boldsymbol{P_i^s}$ is the set of features before including the i-th feature in the s ordering. All in all, the magnitude of the SHAP value ($|\boldsymbol{\phi}_i(f_\kappa, \boldsymbol{x}_t)|$) denotes a variable's contribution to a certain prediction. For instance, a significant negative SHAP value indicates that the variable's value lowers the hygroscopicity estimation, while a significant positive SHAP value denotes a positive increase in the hygroscopicity estimation due to the variable's value. The ML pipeline is shown in Fig. S5.

ML model validation. To find the optimal model for each data set and input variable, CV is performed. To validate the ML proxy for κ, a sensitivity analysis is performed in which models are tested with all possible combinations of input variables (PNSD, chemical composition, and meteorological variables), Table S3. Two different metrics are used to quantify the goodness of fit of the models; the coefficient of determination (R^2) and the root mean square error (RMSE).

$$\hat{y}_t \; or \; \kappa_{ML} \sim \mathrm{f(chem, meteo, PNSD)} \quad (10)$$

$$R^2(\boldsymbol{y}, \hat{\boldsymbol{y}}) = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}$$

$$RMSE(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \sqrt{\frac{1}{T}\sum_t (y_t - \hat{y}_t)^2}$$

Where $\bar{y}$ denotes the average hygroscopicity value. The R^2 is defined as the proportion of variance explained by the model's output. The RMSE is defined as the root of the average square residuals.

The limitations and advantages of the ML method are discussed in the Supplementary Notes 1–3 to this work.

Mixing Index (χ_{proxy}) analysis and performance metrics

Direct quantification of aerosol mixing-state χ requires per-particle mass fractions; such data are rarely available in ambient measurements and are not present in the datasets we have. To characterize aerosol mixing state from available observations, we therefore adopt a dual-proxy framework that uses HTDMA GF spread (σGF) as a physical indicator of particle heterogeneity and bulk ACSM/AMS data as a chemical constraint on total population diversity. Where HTDMA measurements are available, we estimate aerosol mixing state using a hygroscopicity-based proxy (χ_{GF}) derived from the spread of measured GF (Fig. 3A–F). Specifically, we define a normalized mixing index.

$$\chi_{GF} = 1 - \frac{\sigma GF}{\sigma_{GF\,\max}} \quad (11)$$

where σGF is the standard deviation of the growth-factor distribution at a given dry diameter, and σGF_{max} is a site-specific upper bound derived from the observed distribution. This formulation maps large GF variability characteristic of externally mixed aerosol populations with distinct hygroscopic modes to low χ_{GF} and converges to $\chi_{GF} = 1$ for internally mixed populations with narrow growth-factor distributions. Although χ_{GF} does not constitute the rigorous particle-resolved mixing state index χ defined by Riemer and West et al.[63], it provides a physically meaningful observational analogue by directly quantifying inter-particle hygroscopic heterogeneity, consistent with previous HTDMA-based studies, which demonstrate that σGF reflects the degree of mixing[4,27].

Further, we adopt a bulk, chemistry-based proxy (χ_{chem}) when we have chemical composition without GF spread information, motivated by the mixing-state framework of Riemer et al[63]. Using ACSM/AMS-derived mass fractions of the major aerosol components (organics, sulfate, nitrate, ammonium, chloride, and black carbon), we compute the bulk chemical diversity Dγ as the effective number of species based on Shannon entropy[31,63]. This quantity characterizes the total chemical complexity of the aerosol population and represents an upper bound on particle-scale diversity in the absence of size or particle-resolved composition. To enable comparison across sites and conditions (Fig. 3G, H), we define a normalized mixing-state proxy $\chi_{chem} = (D\gamma - 1)/(A\text{-}1)$, where A is the number of chemical components. While $\chi_{AMS/ACSM}$ does not correspond to the rigorous particle-resolved mixing state index χ[63,64], it retains a direct physical interpretation: higher bulk chemical diversity implies a higher likelihood of externally mixed aerosol populations, for which internally mixed bulk closure models are expected to be less representative. Model performance is evaluated using two complementary metrics: a pointwise reduction in absolute hygroscopicity error, Δκ ($[\kappa_{chem}\text{-}\kappa_{measured}] - [\kappa_{ml}\text{-}\kappa_{measured}]$), and the gain in explained variance, $\Delta R^2 = R_{ML}{}^2\text{-}R_{chem}{}^2$ (Fig. S12G, H), both computed relative to measured hygroscopicity. The combined use of hygroscopicity and chemistry-based proxies leverages complementary physical and chemical constraints to approximate aerosol mixing state from observational data, consistent with emerging ML frameworks that infer mixing complexity from bulk aerosol properties[31].

ECHAM simulations and radiative forcing analysis

Sensitivity refers to how a model responds to changes in a certain parameter, typically quantifying how variations in input variables affect key climate outputs such as radiative forcing, temperature, or cloud properties. If sensitivity (ξ) in the aerosol-radiation interactions (RF_{ari}) to hygroscopicity (κ) is known from model simulations, then the change in RF_{ari} due to an updated $\kappa_{estimate}$ can be estimated[65]. To find how the improved aerosol hygroscopicity estimation could potentially influence RF, we calculated the relative change in RF_{direct} when using κ_{ML} instead of the reference κ currently used in the model[12]. As across most of the investigated sites, the empirically estimated TOA radiative forcing (See Text S3 and Figs. 8C and S14) shows an approximately linear dependence on κ across site datasets, the relative change in RF_{direct} (Fig. 5) was obtained using a pre-calculated sensitivity to aerosol hygroscopicity (ξ), based on simulations from the aerosol-climate model ECHAM–HAM (version echam6.1-ham2.2-moz0.978), which is the ECHAM atmospheric general-circulation model coupled to the HAM aerosol module, and the MOZ trace gas chemistry module (disabled for our runs). The model is based on the ECHAM, the HAM aerosol module[66,67] and the MOZ81. Aerosol-radiation interactions are computed using Mie theory[67], where scattering and absorption properties are determined under the assumption of an internally mixed aerosol, accounting for aerosol water content. Given that the refractive index of aerosol water is relatively small (2.0×10^{-7}, by Zhang et al.[68]), absorption enhancement in droplets is considered negligible, meaning changes in the hygroscopicity parameter (κ) primarily impact scattering rather than absorption. The model includes all seven aerosol modes, comprehensively representing aerosol size distributions and variations in wet radii with changing κ values. Consequently, the model effectively captures radiative forcing due to aerosol-radiation interactions (RF_{ari}) and their rapid adjustments. The setup enables the representation of radiative forcing from aerosol-radiation interactions and rapid adjustments. The RF_{ari} estimates

were computed following Ghan et al.[68] using two simulations—one with present-day (PD) emissions and another with preindustrial (PI) emissions, while maintaining a climatological annual cycle of SST. The model was run to obtain RF_{ari} estimates using the reference hygroscopicity values ($\kappa_{inorg} = 0.60$; $\kappa_{org} = 0.06$) and an alternative parameterized κ from[12]. The sensitivity (ξ) was calculated globally by dividing the spatially resolved RF_{ari} simulations by Δκ. The radiative forcing values incorporating regional hygroscopicity were derived by multiplying the sensitivity (ξ) (Figure S8) at each site-specific grid point with Δκ, determined as the difference between the model reference and ML-derived κ as:

$$\Delta RF_{\kappa_{ML}} = \xi(RF, \kappa) * \Delta\kappa_{Ref-ML} \quad (12)$$

The κ_{ML} values from our study generally lead to a substantial negative impact on ΔRF, showing a more cooling effect than the earlier study[12], indicating that the hygroscopicity estimation used in this study captures pronounced effects of aerosols on radiative forcing.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The majority of the data sets on hygroscopicity and aerosol chemical composition used in this study have been previously reported and are available from the original publications cited in the manuscript. Data from the figures in this study (https://doi.org/10.6084/m9.figshare.31901050) and from the Paris campaign at the SIRTA site are available through the AERIS data center (https://across.aeris-data.fr/catalogue/, ACROSS campaign). Also, data are available upon request to the author (deshmukh@tropos.de).

Code availability

The Python library for hygroscopicity estimation and the scripts for all experiments have been published in an open repository (https://codeocean.com/capsule/2440465/tree). Similar codes are available at Bitbucket (https://bitbucket.org/sans-rg/hygroscopicity-estimation).

Received: 24 March 2026; Accepted: 1 April 2026;
Published online: 07 May 2026

References

1. Rosenfeld, D. et al. Global observations of aerosol-cloud-precipitation-climate interactions. *Rev. Geophys.* **52**, 750–808 (2014).
2. Li, Z. et al. Aerosol and monsoon climate interactions over Asia. *Rev. Geophys.* **54**, 866–929 (2016).
3. Climate Change 2021: The Physical Science Basis | Climate Change 2021: The Physical Science Basis. https://www.ipcc.ch/report/ar6/wg1/.
4. Spitieri, C., Gini, M., Gysel-Beer, M. & Eleftheriadis, K. Annual cycle of hygroscopic properties and mixing state of the suburban aerosol in Athens, Greece. *Atmos. Chem. Phys.* **23**, 235–249 (2023).
5. Gysel, M. et al. Closure study between chemical composition and hygroscopic growth of aerosol particles during TORCH2. *Atmos. Chem. Phys*. **7**. https://doi.org/10.5194/acp-7-6131-2007 (2007).
6. Petters, M. D. & Kreidenweis, S. M. A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. *Atmos. Chem. Phys.* **7**, 1961–1971 (2007).
7. Schmale, J. et al. Long-term cloud condensation nuclei number concentration, particle number size distribution and chemical composition measurements at regionally representative observatories. *Atmos. Chem. Phys.* **18**, 2853–2881 (2018).
8. Deshmukh, S. et al. External particle mixing influences hygroscopicity in a sub-urban area. *Atmos. Chem. Phys.* **25**, 741–758 (2025).
9. Paramonov, M. et al. A synthesis of cloud condensation nuclei counter (CCNC) measurements within the EUCAARI network. *Atmos. Chem. Phys.* **15**, 12211–12229 (2015).
10. Li, Y. et al. Nitrogen dominates global atmospheric organic aerosol absorption. *Science (1979)* **387**, 989–995 (2025).
11. Healy, R. M. et al. Predicting hygroscopic growth using single particle chemical composition estimates. *J. Geophys. Res.* **119**, 9567–9577 (2014).
12. Pöhlker, M. L. et al. Global organic and inorganic aerosol hygroscopicity and its effect on radiative forcing. *Nat. Commun*. **14**, 6139 (2023).
13. Riemer, N., Ault, A. P., West, M., Craig, R. L. & Curtis, J. H. Aerosol Mixing State: Measurements, Modeling, and Impacts. *Rev. Geophys.* **57**, 187–249 (2019).
14. Kasparoglu, S., Meskhidze, N. & Petters, M. D. Aerosol mixing state, new particle formation, and cloud droplet number concentration in an urban environment. *Sci. Total Environ.* **951**, 175307 (2024).
15. Wang, X. et al. Direct links between hygroscopicity and mixing state of ambient aerosols: Estimating particle hygroscopicity from their single-particle mass spectra. *Atmos. Chem. Phys.* **20**, 6273–6290 (2020).
16. Titos, G. et al. A global study of hygroscopicity-driven light-scattering enhancement in the context of other in situ aerosol optical properties. *Atmos. Chem. Phys* **21**, 13031–13050 (2021).
17. Enroth, J., Mikkilä, J., Németh, Z., Kulmala, M. & Salma, I. Wintertime hygroscopicity and volatility of ambient urban aerosol particles. *Atmos. Chem. Phys.* **18**, 4533–4548 (2018).
18. Wang, Y. et al. Characterization of aerosol hygroscopicity, mixing state, and CCN activity at a suburban site in the central North China Plain. *Atmos. Chem. Phys.* **18**, 11739–11752 (2018).
19. Li, J. et al. Scattering and absorbing aerosols in the climate system. *Nat. Rev. Earth Environ.* **3**, 363–379 (2022).
20. Vu, T. V., Shi, Z. & Harrison, R. M. Estimation of hygroscopic growth properties of source-related sub-micrometre particle types in a mixed urban aerosol. *NPJ Clim. Atmos. Sci*. **4**, 21 (2021).
21. Fung, P. L. et al. Constructing transferable and interpretable machine learning models for black carbon concentrations. *Environ. Int.* **184**, 108449 (2024).
22. Liu, J. et al. Hygroscopicity of organic aerosols linked to formation mechanisms. *Geophys. Res. Lett.* **48**, e2020GL091683 (2021).
23. Zhang, S. et al. Atmospheric particle hygroscopicity and the influence by oxidation state of organic aerosols in Urban Beijing. *J. Environ. Sci.* **124**, 544–556 (2023).
24. McCoy, D. T. et al. Natural aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo. *Sci. Adv*. **1**, e1500157 (2015).
25. Mandariya, A. K. et al. Measurement report: Hygroscopicity of size-selected aerosol particles in the heavily polluted urban atmosphere of Delhi: impacts of chloride aerosol. *Atmos. Chem. Phys.* **24**, 3627–3647 (2024).
26. Canagaratna, M. R. et al. Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer. *Mass Spectrom. Rev.* **26**, 185–222 (2007).
27. Sjogren, S. et al. Hygroscopicity of the Submicrometer Aerosol at the High-Alpine Site Jungfraujoch, 3580 m a.s.l., Switzerland. *Atmos. Chem. Phys*. **8** www.atmos-chem-phys.net/8/5715/2008/ (2008).
28. Fan, X. et al. Contrasting size-resolved hygroscopicity of fine particles derived by HTDMA and HR-ToF-AMS measurements between summer and winter in Beijing: The impacts of aerosol aging and local emissions. *Atmos. Chem. Phys.* **20**, 915–929 (2020).
29. Wex, H., McFiggans, G., Henning, S. & Stratmann, F. Influence of the external mixing state of atmospheric aerosol on derived CCN number concentrations. *Geophys. Res. Lett.* **37**, 10805 (2010).
30. Gliß, J. et al. AeroCom phase III multi-model evaluation of the aerosol life cycle and optical properties using ground- And space-based remote sensing as well as surface in situ observations. *Atmos. Chem. Phys.* **21**, 87–128 (2021).

31. Zheng, Z. et al. Estimating Submicron Aerosol Mixing State at the Global Scale With Machine Learning and Earth System Modeling. *Earth Space Sci.* **8**, e2020EA001500 (2021).
32. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
33. Ameer, S. et al. Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access* **7**, 128325–128338 (2019).
34. Liu, J., Zhou, D., Zhu, C., Han, J. & Yin, B. Optical and radiative properties of coated black carbon during hygroscopic growth and retrieval errors of mixing state using single-particle soot photometer. *Particuology* **97**, 183–192 (2025).
35. Das, A. et al. Increased particle mass deposition on lung tissue due to industrial and waste-burning activities. *Environ. Int.* **201**, 109548 (2025).
36. Chen, Y. et al. Ammonium chloride associated aerosol liquid water enhances haze in Delhi, India. *Environ. Sci. Technol.* **56**, 7163–7173 (2022).
37. Christodoulou, A. et al. Submicron aerosol pollution in Greater Cairo (Egypt): A new type of urban haze? *Environ. Int*. **186**, 108610 (2024).
38. India is a global warming 'hole,' and scientists aren't sure why. *AAAS Articles DO Group* https://doi.org/10.1126/SCIENCE.ZYVI9VC (2025).
39. DeCarlo, P. F. et al. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. *Anal. Chem.* **78**, 8281–8289 (2006).
40. Weingartner, E. et al. Absorption of light by soot particles: determination of the absorption coefficient by means of aethalometers. *J. Aerosol Sci.* **34**, 1445–1463 (2003).
41. Drinovec, L. et al. The 'dual-spot' Aethalometer: An improved measurement of aerosol black carbon with real-time loading compensation. *Atmos. Meas. Tech*. **8**, 1965–1979 (2015).
42. Massling, A., Leinert, S., Wiedensohler, A. & Covert, D. Hygroscopic Growth of Sub-Micrometer and One-Micrometer Aerosol Particles Measured during ACE-Asia. *Atmos. Chem. Phys*. 7 www.atmos-chem-phys.net/7/3249/2007/ (2007).
43. Ray, A. et al. Seasonal variability in size-resolved hygroscopicity of sub-micron aerosols over the Western Ghats, India: Closure and parameterization. *Sci.e Total Environ.* **869**, 161753 (2023).
44. Wu, Z. J. et al. Relating particle hygroscopicity and CCN activity to chemical composition during the HCCT-2010 field campaign. *Atmos. Chem. Phys.* **13**, 7983–7996 (2013).
45. Denjean, C. et al. Atmospheric Measurement Techniques A new experimental approach to study the hygroscopic and optical properties of aerosols: application to ammonium sulfate particles. *Atmos. Meas. Tech.* **7**, 183–197 (2014).
46. Formenti, P. et al. The aerosols, radiation and clouds in southern africa field campaign in Namibia: overview, illustrative observations, and way forward. *Bull. Am. Meteorol. Soc.* **100**, 1277–1298 (2019).
47. Atabakhsh, S. et al. A 1-year aerosol chemical speciation monitor (ACSM) source analysis of organic aerosol particle contributions from anthropogenic sources after long-range transport at the TROPOS research station Melpitz. *Atmos. Chem. Phys.* **23**, 6963–6988 (2023).
48. Poulain, L. et al. Chemical mass balance of 300 °C non-volatile particles at the tropospheric research site Melpitz, Germany. *Atmos. Chem. Phys.* **14**, 10145–10162 (2014).
49. Shen, C., Zhao, G., Zhao, W., Tian, P. & Zhao, C. Measurement report: Aerosol hygroscopic properties extended to 600 nm in the urban environment. *Atmos. Chem. Phys.* **21**, 1375–1388 (2021).
50. Mallet, M. D. et al. Summertime surface PM1 aerosol composition and size by source region at the Lampedusa island in the central Mediterranean Sea. *Atmos. Chem. Phys.* **19**, 11123–11142 (2019).
51. Atabakhsh, S. et al. Trends of PM1 aerosol chemical composition, carbonaceous aerosol, and source over the last 10 years at Melpitz (Germany). *Atmos. Environ.* **346**, 121075 (2025).
52. Spindler, G. et al. A four-year size-segregated characterization study of particles PM10, PM2.5 and PM1 depending on air mass origin at Melpitz. *Atmos. Environ.* **44**, 164–173 (2010).
53. Huang, S. et al. Aerosol hygroscopicity and its link to chemical composition in a remote marine environment based on three transatlantic measurements. *Environ. Sci. Technol.* **56**, 9613–9622 (2022).
54. Stokes, R. H. & Robinson, R. A. Interactions in aqueous nonelectrolyte solutions. I. Solute-solvent equilibria. *J. Phys. Chem.* **70**, 2126–2131 (1966).
55. Köhler, H. The nucleus in and the growth of hygroscopic droplets. *Trans. Faraday Soc.* **32**, 1152–1161 (1936).
56. Petters, M. D. et al. Towards closing the gap between hygroscopic growth and activation for secondary organic aerosol-Part 2: Theoretical approaches. *Atmos. Chem. Phys.* **9**, 3999–4009 (2009).
57. Wex, H. et al. Atmospheric Chemistry and Physics Towards Closing the Gap between Hygroscopic Growth and Activation for Secondary Organic Aerosol: Part 1-Evidence from Measurements. *Atmos. Chem. Phys*. **9** www.atmos-chem-phys.net/9/3987/2009/ (2009).
58. Gupta, P. et al. Estimating black carbon levels using machine learning models in high-concentration regions. *Sci. Total Environ*. **948**, 174804 (2024).
59. Ferrer-Cid, P. et al. A data-driven framework for air quality sensor networks. *IEEE Internet Things Mag.* **7**, 128–134 (2024).
60. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **13**, 785–794 (2016).
61. Qiu, Y. et al. Predicting atmospheric particle phase state using an explainable machine learning approach based on particle rebound measurements. *Environ. Sci. Technol.* **57**, 15055–15064 (2023).
62. Gao, Z. et al. Explainable machine learning reveals the unknown sources of atmospheric HONO during COVID-19. *ACS EST Air* **1**, 1252–1261 (2024).
63. Riemer, N. & West, M. Quantifying aerosol mixing state with entropy and diversity measures. *Atmos. Chem. Phys.* **13**, 11423–11439 (2013).
64. Jiang, Y., Ma, Y., Zheng, J., Ye, N. & Yuan, C. Characterization of size-resolved aerosol hygroscopicity and liquid water content in Nanjing of the Yangtze River Delta. *J. Environ. Sci.* **151**, 26–41 (2025).
65. Paynter, D. & Frölicher, T. L. Sensitivity of radiative forcing, ocean heat uptake, and climate feedback to changes in anthropogenic greenhouse gases and aerosols. *J. Geophys. Res.: Atmos.* **120**, 9837–9854 (2015).
66. Stier, P. et al. The aerosol-climate model ECHAM5-HAM. *Atmos. Chem. Phys.* **5**, 1125–1156 (2005).
67. Zhang, K. et al. The global aerosol-climate model ECHAM-HAM, version 2: Sensitivity to improvements in process representations. *Atmos. Chem. Phys.* **12**, 8911–8949 (2012).
68. Ghan, S. J. Technical note: Estimating aerosol effects on cloud radiative forcing. *Atmos. Chem. Phys.* **13**, 9971–9974 (2013).
69. Gysel, M., Laborde, M., Olfert, J. S., Subramanian, R. & Gréhn, A. J. Effective density of Aquadag and fullerene soot black carbon reference materials used for SP2 calibration. *Atmos. Meas. Tech.* **4**, 2851–2858 (2011).
70. Kondo, Y. et al. Consistency and traceability of black carbon measurements made by laser-induced incandescence, thermal-optical transmittance, and filter-based photo-absorption techniques. *Aerosol Sci. Technol.* **45**, 295–312 (2011).
71. Park, K., Kittelson, D. B., Zachariah, M. R. & McMurry, P. H. Measurement of inherent material density of nanoparticle agglomerates. *J. Nanopart. Res.* **6**, 267–272 (2004).
72. Alfarra, M. R. et al. A mass spectrometric study of secondary organic aerosols formed from the photooxidation of anthropogenic and biogenic precursors in a reaction chamber. *Atmos. Chem. Phys.* **6**, 5279–5293 (2006).
73. Dinar, E., Mentel, T. F. & Rudich, Y. The density of humic acids and humic like substances (HULIS) from fresh and aged wood burning and pollution aerosol particles. *Atmos. Chem. Phys.* **6**, 5213–5224 (2006).

74. Sullivan, R. C. et al. Effect of chemical mixing state on the hygroscopicity and cloud nucleation properties of calcium mineral dust particles. *Atmos. Chem. Phys.* **9**, 3303–3316 (2009).
75. Zieger, P. et al. Revising the hygroscopicity of inorganic sea salt particles. *Nat. Commun*. **8**, 15883 (2017).
76. Stein-Zweers, D. & V. P. OMI/Aura Multi-wavelength Aerosol Optical Depth. *EarthDATA NASA - Giovanni* https://doi.org/10.5067/Aura/OMI/DATA3004 (2012).

Acknowledgements

PFC, JMB, and JGV work was funded by (grant nos. PID2022-138155OB-I00 and MCIN/AEI/10.13039/501100011033 by ERDF A way of making Europe), Spain, and by (grant nos. 2021SGR-01059 and AGAUR 2023 CLIMA 0097) AGAUR regional projects, Spain. PK acknowledges support from the RECLAIM Network Plus (EP/W034034/1) project. PF and BD acknowledge the AEROCLO-sA project supported by the French National Research Agency (ANR) under (grant no. ANR-15-CE01-0014-01), the French national program LEFE/INSU, the Programme national de Télé-detection Spatiale (PNTS, grant no. PNTS-2016-14), the French National Agency for Space Studies (CNES), and the South African National Research Foundation (NRF) under (grant UID 105958. PEGASUS receives funding as a national facility of the Institut National des Sciences de l'Univers du Centre national de la recherche scientifique (CNRS INSU) within the research infrastructure ACTRIS-FR. MDP acknowledges support from the U.S. Department of Energy, Office of Science, Biological and Environmental Research grant no. DE-SC0021074. A.C. and S.Bwork was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no.856612 (EMME-CARE) and by the STDF-AUF-IRD launched in 2016 (POLCAIR 'Air Pollution in Cairo: sources and impact). The Deutsche Forschungsgemeinschaft (DFG) has supported this research (grant no. WE 2757/4-1). SD, MP, LP, BR, SH, and BW are funded and supported by the DFG. Managed by Leibniz Institute for Tropospheric Research (TROPOS).

Author contributions

S.D. wrote the original manuscript. S.D., A.A., and M.P. conceptualized the study. S.D., P.F.C., and A.A. were responsible for the data management and curation. S.D. and P.F.C. created the figures and tables. P.F.C., J.M.B., J.G.V. and M.V. work on machine learning and analysis. B.R., A.A. and P.F.C. contributed to writing and correcting the manuscript. B.R., S.D. and M.P. work on radiative forcing analysis and interpretation. L.P., S.H. and B.W. contributed to the review and editing of the paper. S.D., B.R., L.P., M.P. and S.H. provided the methodology, acquired funding, and cooperation. S.D., M.D.P., C.D., B.D., A.A., P.F., S.M., G.H. and A.C. provided the observational data for the manuscript from different sites. M.D.P., M.V., C.D., B.D., P.F., J.M.B., Z.W., Sh.H., P.K., S.B., L.P., S.H., B.W., A.A. and M.P. provided supervision and validation. All authors contributed to the scientific discussion, interpretation of the results, and final paper writing.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Diversity, Equity, Ethics, and Inclusion

We are committed to fostering diversity, equity, inclusion, and ethical research practices, ensuring that our work is conducted respectfully for all individuals and communities.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-026-03505-z.

Correspondence and requests for materials should be addressed to Shravan Deshmukh, Ajit Ahlawat or Mira Pöhlker.

Peer review information Communications Earth and Environment thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Alice Drinkwater. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Leibniz Institute for Tropospheric Research, e.V. (TROPOS), Permoserstrasse 15, Leipzig, Germany. [2]Department of Computer Architecture, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. [3]Climate and Atmosphere Research Center (CARE-C), the Cyprus Institute, Nicosia, Cyprus. [4]Météo-France, CNRS, , University of Toulouse, CNRM, Toulouse, France. [5]Aix-Marseille Université, CNRS, LCE, Marseille, France. [6]Université Paris Cité and Univ Paris Est Créteil, CNRS, LISA, Paris, France. [7]Indian Institute of Tropical Meteorology, Pune, India. [8]Indian Institute for Technology New Delhi (IIT-D), New Delhi, India. [9]Global Centre for Clean Air Research (GCARE), School of Engineering, Civil and Environmental Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK. [10]Institute for Sustainability, University of Surrey, Guildford, Surrey, UK. [11]College of Environment and Climate, Jinan University Guangzhou, China. [12]State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University Beijing, China. [13]Spanish Ministry for Ecological Transition, Madrid, Spain. [14]Institute of Environmental Assessment and Water Research, Spanish Research Council (IDAEA-CSIC), Barcelona, Spain. [15]Department of Chemical and Environmental Engineering, University of California Riverside, Riverside, CA, USA. [16]Center for Environmental Research and Technology (CE-CERT), University of California Riverside, Riverside, CA, USA. [17]Department of Geoscience and Remote Sensing, Delft University of Technology (TU Delft), Delft, The Netherlands. [18]Faculty of Physics and Earth Sciences, Leipzig Institute for Meteorology, Leipzig University, Leipzig, Germany. [19]Present address: PSI Center for Energy and Environmental Sciences, Paul Scherrer Institute, Villigen, Switzerland.
✉e-mail: deshmukh@tropos.de; A.S.Ahlawat@tudelft.nl; poehlker@tropos.de