

ARTICLE

OPEN



Estimation of hygroscopic growth properties of source-related sub-micrometre particle types in a mixed urban aerosol

Tuan V. Vu¹, Zongbo Shi^{1,2} and Roy M. Harrison^{1,3}✉

Knowledge of hygroscopic properties is essential to prediction of the role of aerosol in cloud formation and lung deposition. Our objective was to introduce a new approach to classify and predict the hygroscopic growth factors (Gfs) of specific atmospheric sub-micrometre particle types in a mixed aerosol based on measurements of the ensemble hygroscopic growth factors and particle number size distribution (PNSD). Based on a non-linear regression model between aerosol source contributions from PMF applied to the PNSD data set and the measured Gf values (at 90% relative humidity) of ambient aerosols, the estimated mean Gf values for secondary inorganic, mixed secondary, nucleation, urban background, fresh, and aged traffic-generated particle classes at a diameter of 110 nm were found to be 1.51, 1.34, 1.12, 1.33, 1.09 and 1.10, respectively. It is found possible to impute (fill) missing HTDMA data sets using a Random Forest regression on PNSD and meteorological conditions.

npj Climate and Atmospheric Science (2021)4:21; <https://doi.org/10.1038/s41612-021-00175-w>

INTRODUCTION

Processes affecting atmospheric aerosol and its effects on climate change are strongly dependent upon hygroscopic properties^{1,2}. Once a particle is emitted or formed in the atmosphere, it can grow or shrink in size by water vapour uptake due to its hygroscopicity, therefore altering the scattering and absorption of solar radiation and consequently changing the Earth's radiation balance^{3,4}. In addition, hygroscopic properties of aerosols play an important role in determining the impact of aerosols in cloud droplet formation^{5,6}, indicating that particles with a high hygroscopic growth factor (Gf) mode are predominantly scavenged into cloud droplets with an activation efficiency of 57–83%⁶. Particle growth factor is also considered as a vital parameter in determining the deposition of aerosols in the human respiratory system^{2,7,8} as atmospheric aerosols at 200 nm can grow their size to more than double when exposed to a high relative humidity (>99.5%) in the human lung, enhancing their lung deposition efficiencies.

Hygroscopic properties of aerosols can be described by hygroscopic Gf that are commonly measured by a hygroscopicity tandem differential mobility analyser (HTDMA) instrument¹. The Gf is defined as the ratio between the particle diameter measured at a high relative humidity (RH) condition (d_w , RH~90%) and a dry particle diameter measured at a low RH condition (d_p , RH < 10%):

$$Gf(RH) = \frac{d_w(RH)}{d_p(RH < 10\%)} \quad (1)$$

The values of Gf have been widely used to classify hygroscopic properties of aerosols into four groups: nearly-hydrophobic (Gf ~<1.11), less-hygroscopic (Gf ~1.11–1.33), more-hygroscopic (Gf ~1.33–1.85) and sea-salt aerosols (Gf ~>1.85)¹. HTDMA techniques are now used extensively to characterise hygroscopic properties of atmospheric aerosols in different environments, including rural, marine, remote, urban background, and roadside areas². However, there are some disadvantages of HTDMA measurements: (1) a HTDMA instrument can only measure particle

growth factors for several selected diameters in a limited size range (30–350 nm); (2) the measurement does not have high time resolution; (3) possible bias if equilibrium conditions are not adequately reflected⁸; and (4) the effect of condensed vapours on the surface of measured particles creating internally mixed particles and error in the HTDMA measurements^{8–10}.

Particles released from different sources have different chemical composition and sizes that determine their different hygroscopic properties. Vu et al.² reviewed the Gf of particles from three sources, including traffic emissions, biomass burning and nucleation, and found that fresh combustion-generated aerosols are predominantly nearly-hydrophobic or less-hygroscopic particles. During the atmospheric aging process, aerosols can be subjected to other aerosol dynamical processes which may change their shape and chemical composition, leading to a change in their hygroscopic properties. Kotchenruther and Hobbs¹¹ found higher growth factors of aged wood smoke aerosols (1.3–1.5) than those of fresh wood smoke aerosol (1.1–1.3). Aerosols emitted from traffic sources have sometimes been observed to increase their hygroscopic growth factor during atmospheric ageing, but this increment was much lower than that of biomass burning aerosols¹². McMeeking et al.¹³ used a HTDMA coupled with a single-particle soot photometer (HTDMA-SP2) system to measure the refractory black carbon (rBC) aerosol hygroscopicity and found an influence of mixed rBC on the hygroscopic properties of aerosols. Despite a considerable number of HTDMA measurements for ambient aerosols in recent years, there are a few modelling studies on hygroscopic properties of aerosols from sources such as coal combustion or secondary organic aerosols.

Hygroscopic growth is affected by both the inherent hygroscopicity of the material comprising the particle, and for smaller particles (<100 nm) by the influence of the Kelvin Effect upon vapour pressure. This work addresses measured growth factors that are affected by both influences, and does not therefore address directly the inherent hygroscopicity of the particles. However, the growth factor is a practically valuable quantity in the context of

¹Division of Environmental Health & Risk Management, School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, UK. ²Institute of Earth Surface System Science, Tianjin University, Tianjin, China. ³Department of Environmental Sciences/Center of Excellence in Environmental Studies, King Abdulaziz University, PO Box 80203, Jeddah, Saudi Arabia. ✉email: r.m.harrison@bham.ac.uk

actual growth under high humidity conditions and is directly relevant to particle size changes at high relative humidities in both the atmosphere and the human respiratory tract.

The aim of this work was to investigate the relationship between hygroscopic properties of aerosols and their emission sources. To do so, Gf values for particles from different sources have been estimated based on a non-linear regression (NLR) model between the measured Gf of ambient particles and factor contributions obtained by a positive matrix factorisation (PMF) model applied to particle number size distribution (PNSD) data sets. This study also reports the use of a PMF model on a combined data set of PNSD and hygroscopic particle counts for the selected radii to investigate the hygroscopicity of each emission source. Furthermore, this study also develops a new approach for prediction of aerosol growth factors based on their PNSD.

This work has important potential applications. Particle hygroscopicity, as indicated above, is a powerful determinant of the ability of particles to act as cloud condensation nuclei (CCN), and to deposit with high efficiency in the respiratory system. Since measurements of hygroscopic growth factors with the HTDMA do not distinguish the sources of particles, methods such as this, which are capable of associating growth factors with particles from specific sources, are a high priority in understanding the sources of CCN and identifying those particles which are most responsible for contributing to the lung dose of atmospheric pollutants.

RESULTS

Hygroscopic properties of aerosols from measurements

The measured growth factor probability density function (Gf-PDF) presented a bi-modal distribution for each selected particle size of 50, 75, 110, 165, and 265 nm during the two sampling campaigns as shown in Fig. 1. Two dominant fractions of aerosols are found as nearly-hydrophobic ($G_f \sim 1.00\text{--}1.15$, at 90% RH) and more-hygroscopic particles ($G_f \sim 1.35\text{--}1.75$). Larger particles ($d_p > 100$ nm) show greater hygroscopicity by shifting the Gf-probability density towards the mode of greater Gf. There is no significant difference between mean growth factors between summer and winter, but the peak mode of the nearly-hydrophobic and more-hygroscopic groups seems shifted to smaller Gf values in summer which was probably attributable to more atmospheric evaporation of hygroscopic components (e.g. nitrate) of aerosols or possibly more less-hygroscopic secondary organic aerosol (SOA) in summer than winter.

In terms of the fraction of aerosol by number, a greater proportion of nearly-hydrophobic particles (43–45%) were less than 100 nm whereas more-hygroscopic particles were larger than 100 nm. For example, number fractions of aerosols at 265 nm are 35, 4 and 61% for nearly-hydrophobic, less-hygroscopic, and more-hygroscopic aerosol groups. Less-hygroscopic particles were only found to occur appreciably for particles with diameter smaller than 200 nm. In all, 12–15% of such particles were ultrafine particles (diameters <100 nm) that can be attributed to aged combustion-generated aerosols. Sea-salt aerosols were found for some events, but their contribution was not significant (<1%). The number fraction of nearly-hydrophobic particles in this study was greater than those in previous studies in Leipzig (22–31%), Neuherberg (35–42%), Beijing (17–24%), and Shanghai (18–24%)^{14–16}. It is also greater than that measured at a rural site (30–52%), but smaller than that measured at roadside (70%)^{7,17}.

The scheme of model development is shown in Fig. 2. In order to characterise the distribution of growth factors associated with particles from different sources, we developed two approaches: (1) application of a PMF model to the particle size distribution data, followed by use of an NLR model to associate growth factors with

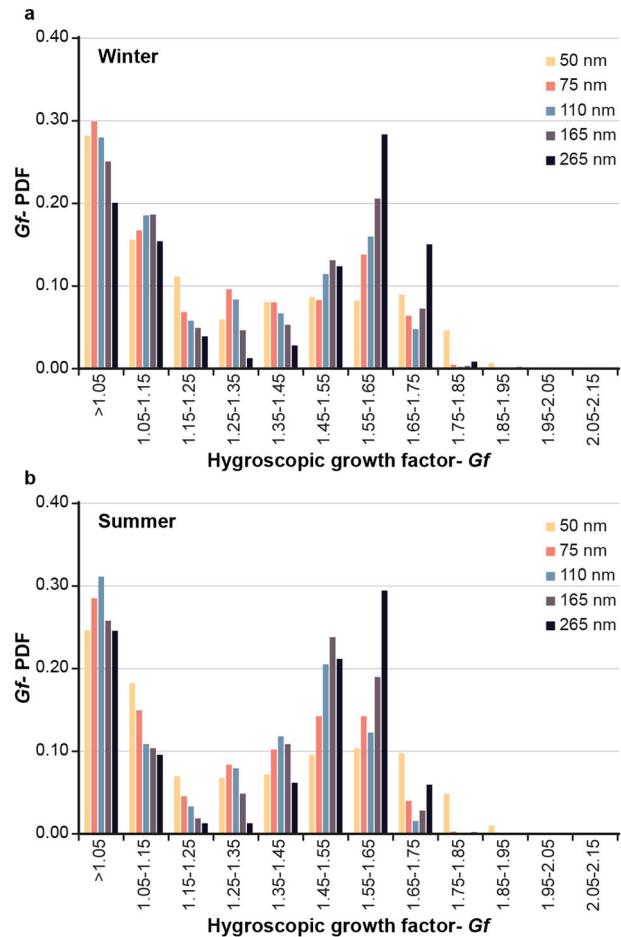


Fig. 1 Hygroscopic growth factor probability density function (Gf-PDF) for particles of 50, 75, 110, 165, and 265 nm. The graphs depict (a) winter, and (b) summer.

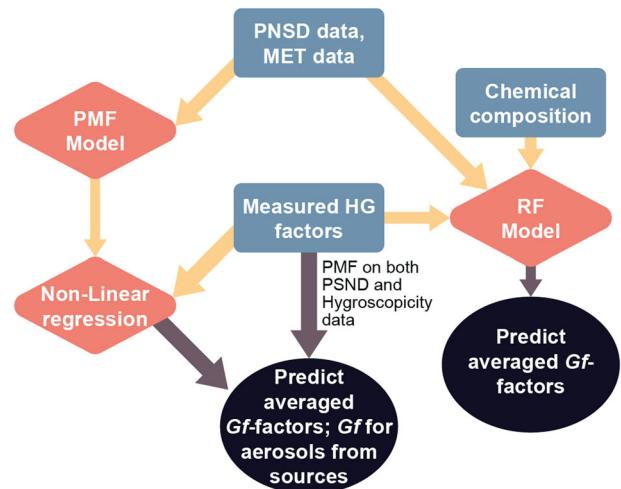


Fig. 2 Schematic of model development scheme. This shows the relationship of different components of the study.

each PMF source-related factor; (2) application of PMF to both the particle size distribution and hygroscopic growth factor data in a single model. Additionally, a Random Forest (RF) model was used to predict the hygroscopic growth factor of ambient aerosols using source contribution factors for missing data imputation for HTDMA data sets.

The optimal solution from the application of PMF to the PNSD data (see "Methods" section) revealed six source-related factors. Figure 3 shows time trend comparisons between the number fraction of nearly-hydrophobic particles and the PMF traffic factor contribution, and between number fraction of more-hygroscopic particles and the PMF secondary and nucleation aerosol contribution. Based on the PMF results obtained from the PMF model on only the PNSD data set, traffic emissions contribute 52% of total particles while secondary aerosols and nucleation contribute 34% of total particles by number which are comparable with 43–45% and 40–44% for nearly-hydrophobic and more-hygroscopic ultrafine particles. It suggests that traffic emissions are the main source of nearly-hydrophobic particles in the urban background atmosphere of London.

Estimated growth factors by non-linear least-squares fitting

Hygroscopic growth factors of particles from different sources obtained by the non-linear least-squares fitting based on Eq. (3) (see "Methods" section) are shown in Table 1¹⁸.

The Gf values of particles from *traffic-generated aerosols* (at 90% RH) are mainly found between 0.92 and 1.29 (except for fresh-traffic particles at a diameter of 265 nm with GF ~1.36) indicating the particles emitted from traffic emissions are nearly hydrophobic or less hygroscopic. These growth factors are in agreement with those reported from chamber experiments and measurements at roadside sites by Weingartner et al.¹⁹, Tritscher et al.¹², Baltenberger et al.¹⁷, Ferron et al.⁷ and Löndahl et al.²⁰. Weingartner et al.¹⁹ reported that the Gf values of aerosols (in the size range of 29–111 nm) released from a four-stroke spark ignition engine ranged from 0.98 to 1.15 at 95% RH.

For particles around 110 nm, there is no significant difference between Gf values for fresh (Gf ~1.09) and aged traffic particles (Gf ~1.10). It is likely that particles in this size range are aggregate or

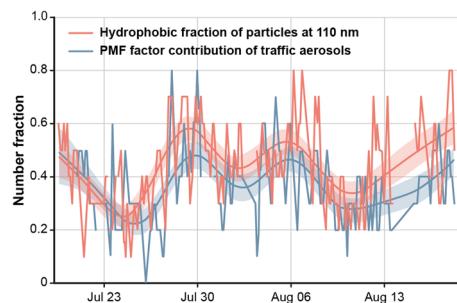


Fig. 3 Time trends in the estimated number fraction. Comparing the measured hydrophobic fraction of particles at 110 nm (red) with the PMF factor contribution of traffic aerosols (blue).

soot particles which change their size slightly during the atmospheric aging process^{12,21}. Aged traffic particles at 50 and 75 nm were found to be more hygroscopic (Gf ~1.19–1.29) than those of fresh traffic emissions (Gf ~0.92–1.20). This may be explained by the evaporation of semi-volatile organic compounds (which are predominantly hydrophobic compounds) from the aerosol surface or the condensation of oxidised species onto the particles during the atmospheric ageing process, creating a more hygroscopic surface. In contrast, larger particles (with diameters at 165 and 265 nm) show higher average Gf values for aerosols from fresh traffic emissions (Gf ~1.26–1.36) than aged traffic emissions ((Gf ~1.11–1.12). Larger fresh traffic-generated aerosols may contain more hygroscopic compounds (i.e. NH₄NO₃) which can evaporate during ageing, shifting to aged traffic aerosols with lower Gf²².

The *urban accumulation particles* are found to be less hygroscopic with the Gf ranging from 1.32 to 1.33 for particles in a range of 75–110 nm, and more hygroscopic (Gf ~1.68) with a diameter of 265 nm. These particles are believed to be a mixture of aged urban combustion-generated aerosols (traffic emissions and wood burning). This source was found to correlate well with both organic compounds ($r^2 \sim 0.77$) and black carbon ($r^2 \sim 0.77$)²³. Aged wood smoke aerosols are known to be more hygroscopic (Gf ~1.3–1.7 at 90–95% RH) compared to fresh wood aerosols^{11,24–26}. A very high growth factor (1.68) for aerosols at 265 nm suggests the urban accumulation mode aerosol contains inorganic salts, many of which have Gf values of 1.6–1.8, or may be mixtures including some compounds with Gf > 2 (at RH 90%)²⁷.

True *nucleation mode particles* fall below the lowest particle size measured by the HTDMA of 50 nm. However, the regression model assigns growth factor values to this PMF source factor which may relate to the upper tail of the distribution. This is an artefact of the PMF, which assigns a small number of larger particles to this factor. Ultrafine particles assigned by PMF to the nucleation factor show growth factors of 1.12–1.30 with higher Gf values for smaller particles. These particles arise from regional nucleation and have grown due to condensation of oxidised species²³. Sakurai et al.²⁸ reported a Gf value of 1.4 for nucleation particles at a diameter of 10 nm from an urban background site in Atlanta. The high Gf factor of nucleation aerosols indicates that these atmospheric nucleation aerosols are probably mainly formed from oxidation of SO₂²⁹ but are also affected by oxidised surface layers. The nucleation factor contributes only a small fraction of particles in the accumulation mode size range.

Mixed *secondary organic and inorganic aerosols* (MIA) are also found to be less hygroscopic (Gf ~1.34–1.48) than secondary inorganic aerosols (SIA) (Gf ~1.46–1.77). These aerosols were predominantly transported from mainland Europe to London during the sampling campaign. The higher Gf value for SIA is due to more hygroscopic compounds contained in SIA than MIA. Vu et al.²³ found that SIA

Table 1. Hygroscopic growth factors of particles from different sources (from both summer and winter samples).

Sources	Gf of particles at different diameters at RH 90%				
	50 nm	75 nm	110 nm	165 nm	265 nm
Urban accumulation mode	1.40 ± 0.04*	1.32 ± 0.02*	1.33 ± 0.02*	1.34 ± 0.04*	1.68 ± 0.15*
Aged traffic emission	1.29 ± 0.02*	1.19 ± 0.03*	1.09 ± 0.06*	1.12 ± 0.07*	1.11 ± 0.07*
Nucleation	1.30 ± 0.44*	1.26 ± 0.21*	1.12 ± 0.26	—	—
Inorganic aerosols	1.77 ± 0.16*	1.66 ± 0.42*	1.51 ± 0.78*	1.51 ± 0.04*	1.46 ± 0.02*
Fresh-traffic emission	1.20 ± 0.04*	0.92 ± 0.32	1.10 ± 0.17*	1.26 ± 0.13*	1.36 ± 0.17*
Mixed inorganic/organic aerosols	—	1.48 ± 0.07*	1.34 ± 0.02*	1.36 ± 0.02*	1.34 ± 0.03*

The uncertainty of Gf is represented by the standard errors which are the square roots of the diagonal elements in the estimated variance–covariance matrix¹⁸. T-test result for testing the hypothesis of zero value for parameter is also provided.

*Significance level (<0.0001; t value Pr(>|t|) ~10E–08).

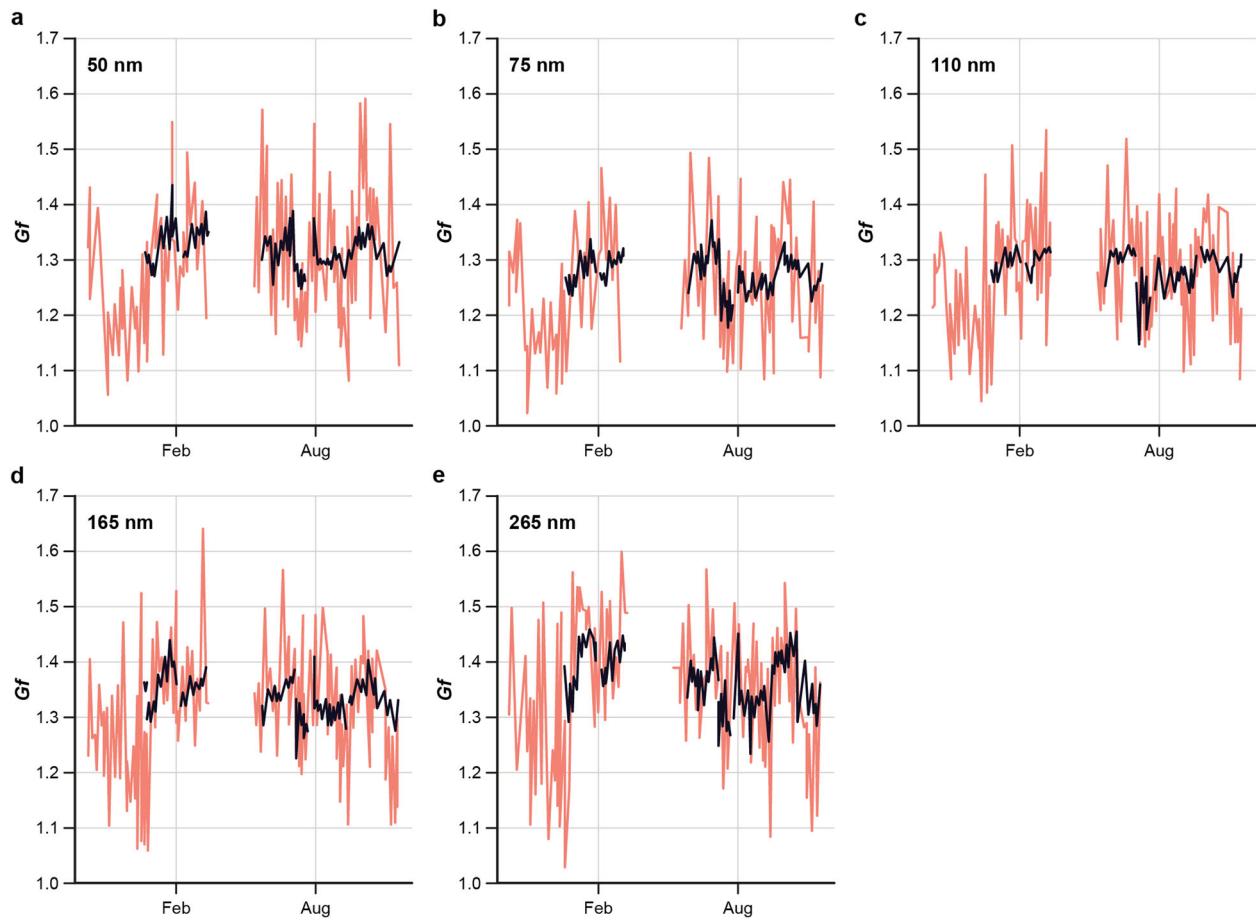


Fig. 4 Comparison of observed growth factor (in red) and modelled factor (in black). Results from the non-linear regression for particles of dry diameter 50, 75, 110, 165 and 265 nm, as indicated.

correlated well with NH_4^+ , NO_3^- and SO_4^{2-} ($r^2 \sim 0.62\text{--}0.80$) whereas MIA has a better relationship ($r^2 \sim 0.63$) with total organic compounds. The Gf values of SIA are in the range of 1.5–1.7 for ammonium sulfate and ammonium nitrate^{30–32}. Hygroscopic growth factors of SOA were reported to be lower with a range of 1.1–1.45 in a previous study³³. The Gf value of MIA is greatly dependent upon the mixing ratio between inorganic salts and SOA³⁴.

Discussion on hygroscopic growth factors from a NLR model

The next step was to associate an estimate of growth factor with each particle class derived from the PMF. Based on averaged growth factors estimated following the NLR model based upon Eq. (3) for aerosols from each source as shown in Table 1¹⁸, and the contribution of each emission source obtained by the PMF model, the hygroscopic growth factors of the mixed ambient aerosol were reconstructed. A comparison between simulated and measured growth factors is shown in Fig. 4.

The simulated growth factors derived from the NLR model could predict well the average temporal trend in the growth factors of ambient aerosols, but failed to predict the lowest and highest points. The poor correlations ($r^2 < 0.4$) between hourly modelled and measured growth factors can be explained by four main reasons:

- (1) The time variation of growth factors of each source: In the NLR model, we used an averaged growth factor value for particles from each source but the Gf values of aerosols from a given source may vary within the sampling time.
- (2) The uncertainty in calculating Gf for a mixture of aerosols (Zdanovskii–Stokes–Robinson (ZSR) mixing rule model)^{34,35}.

- (3) The uncertainty in calculation of number fraction of particles from each source by the PMF model.
- (4) The effects of ambient meteorological conditions on hygroscopic properties of aerosols during ageing processes.

In summary, despite the poor prediction of high time resolution of the aerosol Gf values, the NLR model based on the PMF factors derived from PNSD data sets predicts well the average trend of Gf values and provides estimated Gf values for particles emitted from different sources.

PMF model applied to PNSD and hygroscopic growth factors

In order to further investigate the probability distribution of the Gf values, we applied a PMF model to the data sets combined of PNSD obtained by Scanning Mobility Particle Sizer (SMPS) and hygroscopic particle counts by the HTDMA. The results are shown in Fig. 5. The six source-related factors as seen in the analysis of PNSD data only were identified, but the source profiles are different for urban aerosols which were separated into two groups (less and more-hygroscopic aerosols), and secondary aerosols which was split into two factors (inorganic and mixed secondary) compared to the previous PMF run on the PNSD data only.

For particles from the nucleation source, the hygroscopic growth factor distribution (Gf-PDF) of particles at 50, 75, 110 and 165 nm have a dominant mode of 0.9–1.1. Similar to nucleation particles, the growth factor of particles from fresh traffic and aged traffic sources are mainly distributed at 0.8–1.1 with minor peaks at 1.4 and 1.6. This is as expected for the hydrocarbon and elemental carbon-based traffic exhaust-related particles, but not for the nucleation mode. However, the smallest particle size (50 nm) at which hygroscopicity

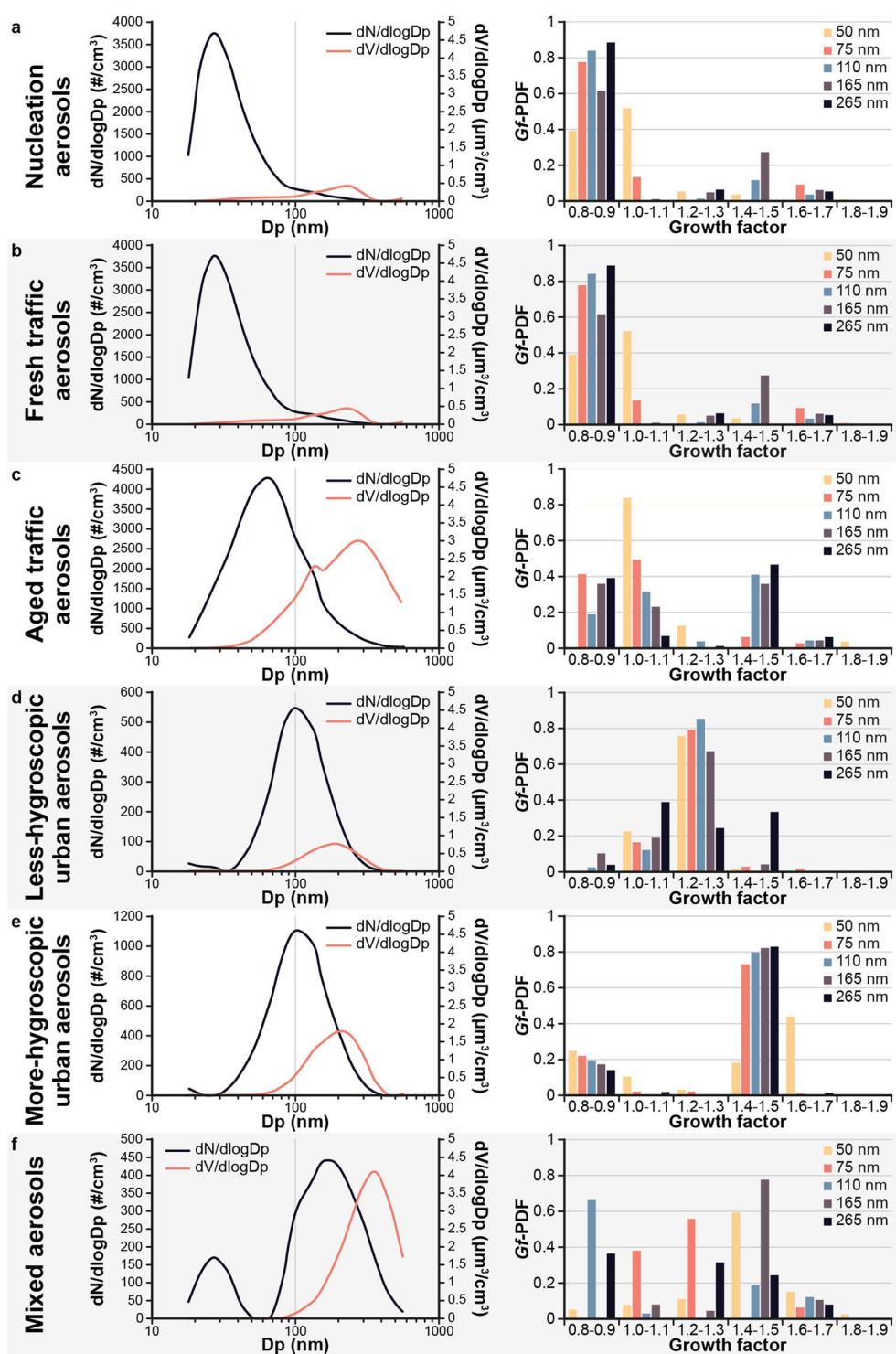


Fig. 5 Results of application of PMF to combined PNSD and hygroscopicity datasets. Source profiles of particle number size distribution and growth factor are shown for each PMF factor.

data are available is almost outside of the size range of the nucleation source particles (Fig. 5) and much larger than the mode. Consequently, some of the particle sizes for which hygroscopicity is estimated are relevant to this particle class. However, there is a very large associated uncertainty as the HTDMA did not measure below 50 nm, and therefore omitted most of this category of particle.

By contrast, particles from the urban accumulation mode and secondary aerosol categories show three dominant Gf modes at

1.0, 1.2 and 1.5 which were assigned into hydrophobic, less-hygroscopic and more-hygroscopic groups. The mixed hygroscopic properties of these particles may be the explanation for the poor fitting of the obtained averaged Gf value from the NLR model, if composition varies with time over the wide range of Gf factors seen in Fig. 5 for SIA and MIA.

Limitations of this approach are that particles from lesser sources (e.g. brake wear or sea salt) may not be separated by the

PMF, and the vast majority of particles deriving from nucleation processes are outside of the measured size range.

RF technique to impute missing data from HTDMA

A machine learning technique, RF, was used to predict growth factors by assuming that the Gf for a given particle size is a function of the factor contribution (g_k) of each source obtained by PMF and meteorological parameters. A part of the data set was used to "train" the model, with the other part used to test its skill (see "Methods" section).

The RF model simulates the hygroscopic growth factors of particles well, as shown in Table 2 and Fig. 6. The model performance does not improve significantly with additional chemical input variables such as NH_4^+ , SO_4^{2-} and black carbon, because these chemical components have a close relationship with the PMF factors.

We tested the application of the RF method in prediction of hygroscopic growth factors with two training methods. In the first training method, we selected a training data set of the first two-third of the whole data set, and then used the rest of the data in August as the testing data set. The model performed very poorly ($r < 0.5$) in predicting the hygroscopic growth factors of particles for the new period. The poor performance may be caused by the difference of hygroscopic growth factors of particles between winter and summer, leading to an inadequate training data set. It is out of the scope of our study to investigate this.

In the second model training method, we randomly select the training data sets (accounting for 70% of the whole data sets)

through the whole sampling period. The model performance was found much better as shown in the Supplementary Information, indicating that we can use this method to impute (fill) the missing data from HTDMA measurements. This is however dependent upon the use of a training data set.

DISCUSSION

This study has introduced and investigated new approaches for estimation of hygroscopic properties of sub-micrometre aerosols using the PNSD. As a result, both an NLR model and a PMF model applied to PNSD and hygroscopicity data sets can predict well the hygroscopic growth factors of particles emitted from different sources without knowledge of chemical composition from each source. In addition, the PMF model can show not only the averaged value but also a probability distribution function of each source's growth factors. This study confirms that aerosol from SIA and MIA sources was found to be more hygroscopic whereas combustion creates dominant fractions of nearly-hydrophobic and less-hygroscopic groups. Furthermore, it found that a NLR model could predict well the average trends of ambient aerosols, but it fails to predict the high time resolution ambient Gf values. This study introduced an RF technique which shows a good capability to predict hygroscopic growth factors of aerosols based on the source factors and ambient meteorological conditions. It performs well with high time resolution data, resulting in the application of this technique in the imputation of missing data for the data sets obtained by a HTDMA system.

METHODS

Measurements and data sources

PNSD, hygroscopic properties and other air pollutant concentrations were measured by the University of Manchester at a monitoring station located in the grounds of Sion Manning School in North Kensington, which is representative of a typical urban background area of London, UK³⁶. Those measurements were conducted during two intensive sampling campaigns (January–February and July–August 2012) within the Clean Air for London (ClearLo) project³⁷.

PNSDs were measured by an SMPS system which consisted of an electrostatic classifier (EC, TSI model 3080) and a condensation particle

Table 2. Quality of fit of RF model for test data sets.

dp (nm)	r^2	RMSE
50	0.82	0.0094
75	0.75	0.0060
110	0.72	0.0047
165	0.83	0.0056
265	0.70	0.0073

r correlation coefficient, RMSE root mean square error.

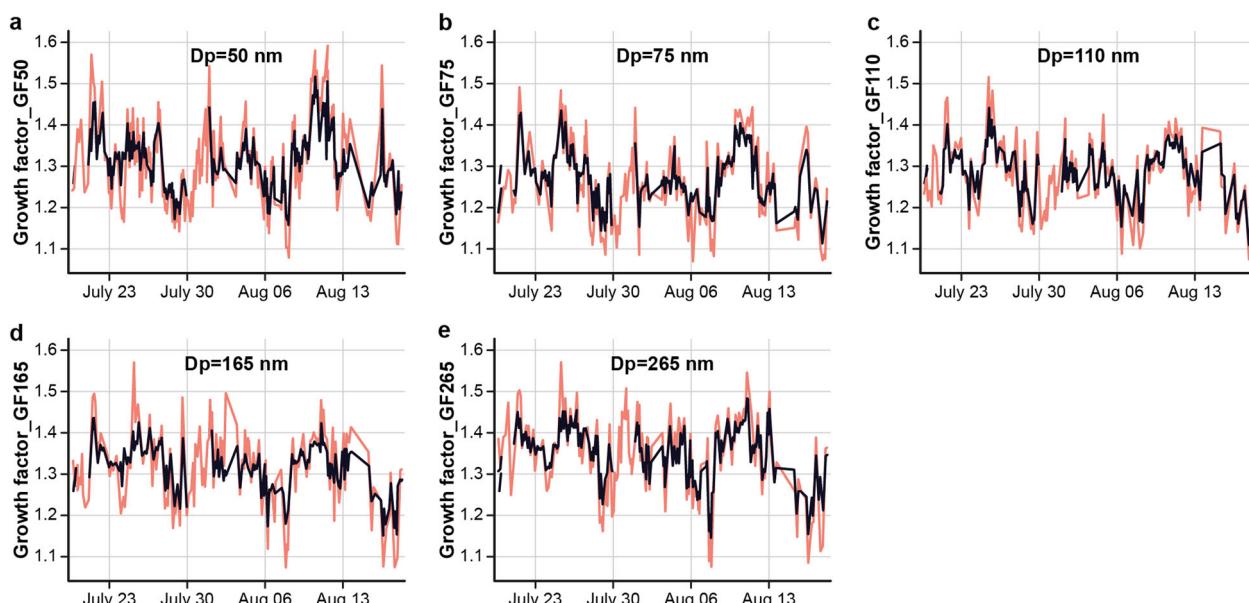


Fig. 6 Comparison of time series of observed growth factors (in red) with those modelled using Random Forest (in black). Each graph relates to a particle size, as indicated.

description of the PMF method and its results (Supplementary Figs. 1 and 2) appear in the Supplementary Information.

A decision tree model based on RF algorithm

A decision tree model based on an RF algorithm has been used recently for classification and regression of time series data sets^{44,45}. The RF model is an ensemble model consisting of hundreds of models (referred as a decision tree with different rules) to give the final decision. This model uses a small amount of training data by the bootstrap technique, and the final model is the average of all sets of predictions. Firstly, the algorithm (referred as bagging) will select the predictor variables randomly with replacement from the training set and select the best predictors out of the random samples to partition the data. A set derived from bagging (referred as out-of-bag data) grow a single tree. The final prediction is the mean of predictions produced and aggregated from different single trees.

In investigating the relationship of Gf and the source contributions to aerosol, the Gf for a given particle size assumed to be a function of the factor contribution (g_k) of each source obtained by PMF and meteorological parameters as in the following equation:

$$Gf \sim f(g_k, wd, ws, RH, temp) \quad (5)$$

PNSD can also be the input instead of g_k , but use of many variables could cause a model over-fitting problem. To investigate the effects of chemical composition, we also added concentrations of ionic species (SO_4^{2-} , NO_3^- , Cl^- , NH_4^+) and black carbon into the model as input variables. The RF code was developed based on the "random forest" function from the "Keras library" in Python (version 2.3) or the "H2O2" package in R. The data set was split by a fraction of 0.7 to train the model and 0.3 for testing the model. Optimised tuning parameters for an RF model are: number of variables randomly sampled is 3; number of trees is 100; and minimum size of terminal nodes is 3.

In our RF method, the relationships between independent variable (hygroscopic growth factor of particle at a certain dry diameter of 50, 75, 110, 165 and 265) and its predictor features including the factor contribution (g_k , $k = 1:6$) of each source obtained by PMF and meteorological parameters (wind speed, wind direction, temperature and humidity) are built based on decision trees. The performance of the model was evaluated by RMSE and r^2 values. The good goodness of fitting as shown in Supplementary Table 1 indicated that RF could be used to impute the missing data sets from the HTDMA based on the PNSD and weather data.

DATA AVAILABILITY

Data supporting this publication are openly available from the UBIRA eData repository at <https://doi.org/10.25500/edata.bham.00000390>.

Received: 23 October 2019; Accepted: 24 February 2021;
Published online: 25 March 2021

REFERENCES

1. Swietlicki, E. et al. Hygroscopic properties of submicrometer atmospheric aerosol particles measured with H-TDMA instruments in various environments—a review. *Tellus B Chem. Phys. Meteorol.* **60**, 432–469 (2008).
2. Vu, T. V., Delgado-Saborit, J. M. & Harrison, R. M. A review of hygroscopic growth factors of submicron aerosols from different sources and its implication for calculation of lung deposition efficiency of ambient aerosols. *Air Qual. Atmos. Health* **8**, 429–440 (2015).
3. McInnes, L., Bergin, M., Ogren, J. & Schwartz, S. Apportionment of light scattering and hygroscopic growth to aerosol composition. *Geophys. Res. Lett.* **25**, 513–516 (1998).
4. Pan, X. L. et al. Observational study of influence of aerosol hygroscopic growth on scattering coefficient over rural area near Beijing mega-city. *Atmos. Chem. Phys.* **9**, 7519–7530 (2009).
5. Reutter, P. et al. Aerosol- and updraft-limited regimes of cloud droplet formation: influence of particle number, size and hygroscopicity on the activation of cloud condensation nuclei (CCN). *Atmos. Chem. Phys.* **9**, 7067–7080 (2009).
6. Väistönen, O. et al. In-cloud measurements highlight the role of aerosol hygroscopicity in cloud droplet formation. *Atmos. Chem. Phys.* **16**, 10385–10398 (2016).
7. Ferron, G. A., Karg, E., Busch, B. & Heyder, J. Ambient particles at an urban, semi-urban and rural site in Central Europe: hygroscopic properties. *Atmos. Environ.* **39**, 343–352 (2005).

counter (TSI model 3775). The instrument was set up to collect PNSDs every 15 min with six scans of 2.5 min covering the size range 14.6–623 nm. The size distribution is measured after the aerosol has passed through a dryer, following the guidelines set out by Wiedensohler et al.³⁸. The hygroscopic growth factors and number fraction of sub-micrometre aerosols were measured by a HTDMA system at $90 \pm 0.8\%$ RH with five selected initial dry sizes at diameters of 50, 75, 110, 165 and 265 nm operated by The University of Manchester. The scan time of HTDMA was 15 min for each run. Both SMPS and HTDMA measurements were conducted in a same location and sampling time period. The Gf-PDF was then retrieved using a TDMAinv inversion³⁸. PNSD and hygroscopic data sets were extracted from the Centre for Environmental Data Analysis (CEDA database, <https://catalogue.ceda.ac.uk/uuid/03cf72a33d1fcf00908bf9eca3be7eca>) and have been described by Whitehead et al.¹⁰.

Other hourly data sets on meteorological parameters, aerosol chemical composition data (ion species and black carbon) were extracted from the Department for Environment Food & Rural Affairs website (DEFRA, <http://uk-air.defra.gov.uk/>). All data were averaged on an hourly or three-hourly (for Gf) basis.

First approach using PMF and an NLR model

The application of PMF models to PNSD data has previously been used for source apportionment of particles by number^{39,40}. By this approach, each size bin in the PNSD data set is considered as an input variable. To reduce the uncertainty which is associated with the SMPS measurements, three consecutive original SMPS size bins were summed into one new size bin. A profile of 1170 hourly PNSD data sets (26/01–11/02/2012 and 21/07–23/08/2012) comprising 32 new size bins ranging from 15 to 500 nm were input in the US EPA PMF model version 5. The PMF model can reduce the dimensions of PNSD data sets by identifying the number of factors (p), the size profile (f) of each source, and the amount of number (g) contributed by each factor to each individual measurement using the following equation:

$$x_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (2)$$

where x_{ij} is the particle number concentration of size bin j on the i th sample and e_{ij} is the residual for the sample.

The details of this PMF method have been described in our previous study. This reported six identified sources: urban accumulation mode, nucleation, SIA, MIA, fresh and aged traffic emissions in the London samples. In the current study, we utilised source contributions and profiles obtained from the PMF model to investigate the influence of sources upon the hygroscopic properties of aerosols.

It was assumed that the average growth factor of mixed aerosols can be estimated based on the ZSR mixing rule^{41,42}.

$$Gf_{\text{mixed}} = \left(\sum_k \varepsilon_k Gf_k^3 \right)^{1/3} \quad (3)$$

Gf_{mixed} is averaged growth factor which is archived from the HTDMA system after using the TDMAinv fitting⁴³. Gf_k is an averaged growth factor of aerosols from each source and ε_k is the number fraction of particle from each source (k) which is estimated from the PMF factor contribution and factor profile. Therefore, the number fraction of particles at a given diameter in a size bin (j) from a source (k) for the i th sample is calculated by the following equation:

$$\varepsilon_k = g_{ik} f_{kj} / \sum_{k=1}^p g_{ik} f_{kj} \quad (4)$$

The Gf_k values were obtained by a non-linear least-squares fitting for the Eq. (3). The "nls" package in R (or "scipy.optimize" library in Python version 2.3) was used for our calculations. To reduce errors arising from the time lag between the SMPS and HTDMA instruments, hourly values of measured Gf, and PMF factor contribution were averaged into three-hourly data before running a non-linear least-squares fitting.

Use of combined PNSD and growth factor data in a PMF model

In this approach, both PNSD data and hygroscopic growth factor data were used as inputs to the PMF model. In a single step, this method gives an output of source-related factors, for each of which there is a particle size distribution and a set of Gf-PDFs for five selected particle sizes. A brief

8. Good, N. et al. Widening the gap between measurement and modelling of secondary organic aerosol properties? *Atmos. Chem. Phys.* **10**, 2577–2593 (2010).
9. Topping, D. & McFiggans, G. Tight coupling of particle size, number and composition in atmospheric cloud droplet activation. *Atmos. Chem. Phys.* **12**, 3253–3260 (2012).
10. Whitehead, J. D., Irwin, M., Allan, J. D., Good, N. & McFiggans, G. A meta-analysis of particle water uptake reconciliation studies. *Atmos. Chem. Phys.* **14**, 11833–11841 (2014).
11. Kotchenruther, R. A. & Hobbs, P. V. Humidification factors of aerosols from biomass burning in Brazil. *J. Geophys. Res. Atmospheres* **103**, 32081–32089 (1998).
12. Tritscher, T. et al. Changes of hygroscopicity and morphology during ageing of diesel soot. *Environ. Res. Lett.* **6**, 034026 (2011).
13. McMeeking, G. R., Good, N., Petters, M. D., McFiggans, G. & Coe, H. Influences on the fraction of hydrophobic and hydrophilic black carbon in the atmosphere. *Atmos. Chem. Phys.* **11**, 5099–5112 (2011).
14. Massling, A. et al. Size segregated water uptake of the urban submicrometer aerosol in Beijing. *Atmos. Environ.* **43**, 1578–1589 (2009).
15. Massling, A., Stock, M. & Wiedensohler, A. Diurnal, weekly, and seasonal variation of hygroscopic properties of submicrometer urban aerosol particles. *Atmos. Environ.* **39**, 3911–3922 (2005).
16. Tschiersch, J., Busch, B. & Fogh, C. L. Measurements of concentration, size distribution and hygroscopicity of Munich winter aerosol. *J. Aerosol Sci.* **28**, S209–S210 (1997).
17. Baltensperger, U. et al. Urban and rural aerosol characterization of summer smog events during the PIPAPO field campaign in Milan, Italy. *J. Geophys. Res. Atmos.* **107**, 8193 (2002). D22.
18. Ritz, C. & Streibig, J. (Eds.), in *Nonlinear Regression with R (Use R)* (Springer, 2008).
19. Weingartner, E., Burtscher, H. & Baltensperger, U. Hygroscopic behavior of combustion aerosols. *J. Aerosol Sci.* **24**, S371–S372 (1993).
20. Löndahl, J. et al. Experimentally determined human respiratory tract deposition of airborne particles at a busy street. *Environ. Sci. Technol.* **43**, 4659–4664 (2009).
21. Dua, S. K., Hopke, P. K. & Raunemaa, T. Hygroscopicity of diesel aerosols. *Water Air Soil Pollut.* **112**, 247–257 (1999).
22. Wittbom, C. et al. Effect of solubility limitation on hygroscopic growth and cloud drop activation of SOA particles produced from traffic exhausts. *J. Atmos. Chem.* **D20**, 4638 (2018).
23. Vu, T. V., Beddows, D. C. S., Delgado-Saborit, J. M. & Harrison, R. M. Source apportionment of the lung dose of ambient submicrometre particulate matter. *Aerosol Air Qual. Res.* **16**, 1548–1557 (2016).
24. Decesari, S. et al. Water soluble organic compounds formed by oxidation of soot. *Atmos. Environ.* **36**, 1827–1832 (2002).
25. Martin, M. et al. Hygroscopic properties of fresh and aged wood burning particles. *J. Aerosol Sci.* **56**, 15–29 (2013).
26. Shingler, T. et al. Airborne characterization of subsaturated aerosol hygroscopicity and dry refractive index from the surface to 6.5 km during the SEAC4RS campaign. *J. Geophys. Res. Atmos.* **121**, 4188–4210 (2016).
27. Koehler, K. A. et al. Water activity and activation diameters from hygroscopicity data —Part II: Application to organic species. *Atmos. Chem. Phys.* **6**, 795–809 (2006).
28. Sakurai, H. et al. Hygroscopicity and volatility of 4–10 nm particles during summertime atmospheric nucleation events in urban Atlanta. *J. Geophys. Res. Atmos.* **110**, D22S04 (2005).
29. Jung, J., Adams, P. J. & Pandis, S. N. Simulating the size distribution and chemical composition of ultrafine particles during nucleation events. *Atmos. Environ.* **40**, 2248–2259 (2006).
30. Gysel, M., Weingartner, E. & Baltensperger, U. Hygroscopicity of aerosol particles at low temperatures. 2. Theoretical and experimental hygroscopic properties of laboratory generated aerosols. *Environ. Sci. Technol.* **36**, 63–68 (2002).
31. Wise, M. E., Surratt, J. D., Curtis, D. B., Shilling, J. E. & Tolbert, M. A. Hygroscopic growth of ammonium sulfate/dicarboxylic acids. *J. Geophys. Res. Atmos.* **108**, 4638 (2003).
32. Müller, A. et al. Effects of chemical composition and mixing state on size-resolved hygroscopicity and cloud condensation nuclei activity of submicron aerosols at a suburban site in northern Japan in summer. *J. Geophys. Res. Atmos.* **122**, 9301–9318 (2017).
33. Duplissy, J. et al. Relating hygroscopicity and composition of organic aerosol particulate matter. *Atmos. Chem. Phys.* **11**, 1155–1165 (2011).
34. Peng, C., Jing, B., Guo, Y.-C., Zhang, Y.-H. & Ge, M.-F. Hygroscopic behavior of multicomponent aerosols involving NaCl and dicarboxylic acids. *J. Phys. Chem. A* **120**, 1029–1038 (2016).
35. Liu, Q., Jing, B., Peng, C., Tong, S., Wang, W. & Ge, M. Hygroscopicity of internally mixed multi-component aerosol particles of atmospheric relevance. *Atmos. Environ.* **125**, 69–77 (2016).
36. Bigi, A. & Harrison, R. M. Analysis of the air pollution climate at a central urban background site. *Atmos. Environ.* **44**, 2004–2012 (2010).
37. Bohnenstengel, S. I. et al. Meteorology, air quality, and health in London: the ClearFlo project. *Bull. Am. Meteorol. Soc.* **96**, 779–804 (2014).
38. Wiedensohler, A. et al. Mobility particle size spectrometers: harmonization of technical standards and data structure to facilitate high quality long-term observations of atmospheric particle number size distributions. *Atmos. Meas. Tech.* **5**, 657–685 (2012).
39. Vu, T. V., Delgado-Saborit, J. M. & Harrison, R. M. Review: Particle number size distributions from seven major sources and implications for source apportionment studies. *Atmos. Environ.* **122**, 114–132 (2015b).
40. Beddows, D. C. S., Harrison, R. M., Green, D. & Fuller, G. Receptor modelling of both particle composition and size distribution data from a background site in London UK. *Atmos. Chem. Phys.* **15**, 10107–10125 (2015).
41. Stokes, R. H. & Robinson, R. A. Interactions in aqueous nonelectrolyte solutions. I. Solute-solvent equilibria. *J. Phys. Chem.* **70**, 2126–2131 (1966).
42. Svenningsson, B. et al. Hygroscopic growth and critical supersaturations for mixed aerosol particles of inorganic and organic compounds of atmospheric relevance. *Atmos. Chem. Phys.* **6**, 1937–1952 (2006).
43. Gysel, M., McFiggans, G. B. & Coe, H. Inversion of Tandem Differential Mobility Analyser (TDMA) measurements. *J. Aerosol Sci.* **40**, 134–151 (2009).
44. Carslaw, D. C. & Taylor, P. J. Analysis of air pollution data at a mixed source location using boosted regression trees. *Atmos. Environ.* **43**, 3563–3570 (2009).
45. Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E. & Hueglin, C. Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmos. Chem. Phys.* **18**, 6223–6239 (2018).

ACKNOWLEDGEMENTS

This research was supported by the UK Natural Environment Research Council funding through the AIRPOLL-Beijing project within the APHH programme (NE/N007190/1). We would like to thank the University of Manchester who collected the initial data set as part of the ClearFlo project (NERC funded under Grant NE/H00324X/1) for data sets archived on CEDA.

AUTHOR CONTRIBUTIONS

T.V.V. conceived the study and carried out the calculations. R.M.H. and Z.S. provided critical comments, drafted parts of the manuscript and revised earlier versions of the text.

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/s41612-021-00175-w>.

Correspondence and requests for materials should be addressed to R.M.H.

Reprints and permission 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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.