

Enhancing Differentiation of Oxygenated Organic Aerosol: A Machine Learning Approach to Distinguish Local and Transboundary Pollution

Lu Lei, Wei Xu,* Chunshui Lin, Baihua Chen, Kirsten N. Fossum, Darius Ceburnis, Colin O'Dowd, and Jurgita Ovadnevaite*




Cite This: *ACS EST Air* 2025, 2, 891–902



Read Online

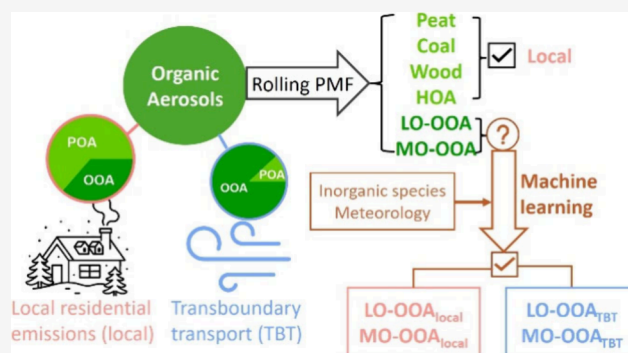
ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: Accurate source apportionment of particulate matter (PM), especially of organic aerosol (OA), is crucial for targeted mitigation efforts. Positive Matrix Factorization (PMF) is powerful in source attribution of primary OA (POA); however, it often struggles to differentiate sources of oxygenated OA (OOA) due to their similar chemical profiles. In this study, a support vector regression machine learning (ML) model was developed to enhance the OOA source apportionment in Dublin from 2016 to 2023. Rolling PMF analysis identified four POA factors and differentiated OOA into less- and more-oxidized (LO-OOA and MO-OOA), highlighting the significant role of the OOA (47–74% of total OA). The ML model further distinguished locally produced OOA (LO-OOA_{local} and MO-OOA_{local}) from transboundary transport OOA and exhibited robust performance across different pollution scenarios. The relative importance analysis revealed that LO-OOA_{local} was more impacted by fossil fuel emissions like hydrocarbon-like OA (20%) and coal (14%), whereas MO-OOA_{local} was most influenced by LO-OOA (17%), providing insights into their sources and formation mechanisms. During a mixed pollution episode, the results show that despite the significant contribution of transboundary transport, local heating emissions were more critical sources of OA, with local OA accounting for 68% of total OA and reaching 78% during heating hours. These findings highlight the ongoing need to reduce local emissions to achieve cleaner air in Dublin. The ML model's ability to quantitatively separate local and transboundary OOA offers invaluable insights for future air quality regulations.



KEYWORDS: *Oxygenated organic aerosol, Source apportionment, Rolling PMF, Machine learning, Urban air pollution*

1. INTRODUCTION

Particulate matter (PM) suspended in the atmosphere significantly impairs air quality and influences the regional and global climate directly and indirectly, introducing large uncertainties in radiative forcing estimation.^{1,2} Additionally, PM adversely impacts human health, causing millions of premature deaths every year.^{3–5} Organic aerosol (OA), constituting 20–90% of submicron PM,⁶ has been found to be more toxic than inorganic species such as nitrate and sulfate, raising greater health concerns.^{7–9} However, OA remains the least understood component due to the high complexity in its composition, sources, physicochemical properties, and formation pathways.¹⁰ Positive Matrix Factorization (PMF) has been widely used as a powerful tool achieving meaningful OA source apportionment for better understanding of its origins.^{6,11,12} While PMF analysis can provide valuable information on primary OA (POA) sources, especially when combined with the multilinear engine algorithm (ME-2), it faces significant challenges in distinguishing oxygenated OA

(OOA) from different sources/processes. This is mainly because OOA loses original source signatures as it undergoes atmospheric oxidation, leading to increasingly similar chemical profiles.¹³ When measured by aerosol mass spectrometer, OOA tends to fragment into only a few main ions, such as CO₂⁺ and C₂H₃O⁺,^{6,14} posing challenges for source apportionment, particularly for unit mass resolution data. As a result, OOA factors are typically classified by their relative oxidation degree, e.g., less oxidized OOA (LO-OOA), which serves as a surrogate of fresher, less-aged secondary OA (SOA) and more oxidized OOA (MO-OOA), a proxy of regional, more-aged SOA.^{11,15} However, this differentiation provides limited

Received: November 26, 2024

Revised: April 5, 2025

Accepted: April 7, 2025

Published: April 15, 2025



information on their sources and formation processes.¹² Additionally, the PMF model has inherent limitations. Most importantly, it assumes linear relationships between sources and observations,¹⁶ which may not always be true for complex atmospheric processes where nonlinear interactions and secondary formation are involved, introducing uncertainty and ambiguity in OA source apportionment.

In recent decades, machine learning (ML) algorithms have been widely applied in atmospheric science across various topics.^{17,18} For example, extensive studies have utilized machine learning methods on air quality forecasting and prediction,^{19,20} evaluating the relative influences on PM concentrations from emissions and meteorological conditions,^{21,22} investigating the source impacts on visibility,²³ and exploring the nonlinear relationship among volatile organic compounds (VOCs), PM_{2.5}, and O₃.²⁴ Additionally, the applications of machine learning algorithms for PM source apportionment are also rapidly increasing and are less computationally expensive. For instance, Qiao et al.²⁵ utilized the decision tree algorithm to attribute atmospheric oxygenated organic molecules to their precursors. Heikkinen et al.²⁶ deployed the *k*-means clustering method to identify OA subtypes in a remote boreal forest site in Southern Finland, where the OA mass concentration remains low and traditional PMF is ineffective. Pande et al.²⁷ developed a two-step machine learning method, combining a multinomial logistic classifier and an ensemble regression model for rapid OA source apportionment. The results showed high classification accuracy and broad qualitative agreement with PMF analysis. These extensive and successful applications of machine learning algorithms on air pollution data analysis demonstrate their unique advantages in capturing nonlinear effects and discerning subtle patterns within multidimensional data sets. This makes machine learning methods well-suited to enhance the apportionment of the OOA source, linking the OOA to more specific sources.

Ireland once experienced extreme air pollution in the 1980s primarily caused by coal combustion.²⁸ Although a series of bans on smoky coal burning since 1990 have significantly alleviated the air pollution in Ireland, Lin et al.²⁹ pointed out that extreme air pollution events were still regularly occurring in Dublin during the residential heating season. The OA source apportionment revealed that these extreme air pollution events primarily result from disproportionately high emissions from domestic solid fuel burning within the broader Dublin urban area (referred to as local). Importantly, in addition to significantly elevated primary species, OOA also shows substantial increase during these local events.³⁰ On the other hand, air pollution events resulting from transboundary transport are also often observed in Ireland, with OA being overwhelmingly dominated by OOA.^{31,32} In other words, the OOA always plays a crucial role in causing air pollution in Dublin. However, as highlighted above, traditional PMF analysis is often insufficient for OOA attribution, leading to significant ambiguity in its source identification and quantification of relative contributions, which hinders the implementation of more targeted control measures to reduce OA pollution.

In this study, the capabilities of machine learning in enhancing source apportionment of OOA were explored. A supervised machine learning model was developed in combination with rolling-PMF analysis, trained with a carefully selected data set from pollution events dominated by local

emissions. Model performance was evaluated using statistical metrics and further validated with two distinct pollution episodes: one dominated by local sources and one dominated by transboundary sources. The ML model was also applied to a mixed pollution event with overlapping impacts from local and transboundary sources to quantify their relative contributions on the OOA. Importantly, none of the three pollution episodes selected for case studies were included in the training data set, ensuring valid evaluations.

2. MATERIALS AND METHODS

2.1. Instruments and Data Collection. The measurements of chemical-specified submicrometer PM (i.e., PM₁, particulate matter with a diameter less than 1 μm) were conducted at the urban background site in Dublin, Ireland, from August 2016 to December 2023. Detailed introduction of the sampling site can be found in Lin et al.³⁰ An Aerodyne Quadrupole Aerosol Chemical Speciation Monitor (Q-ACSM, Aerodyne Research Inc., USA) was deployed for the measurements of nonrefractory species in PM₁ (NR-PM₁) including Organic Aerosol (OA), sulfate (SO₄), nitrate (NO₃), ammonium (NH₄), and chloride (Cl). More details about the sampling protocol of Q-ACSM can be found in previous studies.^{33,34} The Q-ACSM deployed in this study was regularly calibrated following standard calibration protocols to ensure data quality.³⁴ Additionally, the chemical composition collection efficiency³⁵ was applied to NR-PM₁ species consistently across all years. More details on NR-PM₁ measurements in Dublin are summarized in S1.1, including deployed Q-ACSM instruments, data coverage, NR-PM₁ concentration ranges (Tables S1–S2), and also instrument stability (Figure S1a). An Aethalometer (model AE33 from Magee Scientific) was also deployed to measure the optical attenuation at 7 wavelengths. The mass concentration of equivalent black carbon (eBC) was then retrieved using attenuation at 880 nm with a standard mass absorption cross-section value of 7.77 m² g⁻¹. In addition, a Scanning Mobility Particle Sizer (SMPS) was collocated to measure the PM number concentration and size distribution in the range 10–500 nm. Hourly meteorological parameters, including wind speed (WS), wind direction (WD), relative humidity (RH), and ambient temperature (*T*) were from Dublin Airport (<https://www.met.ie/climate/available-data/historical-data>). To ensure consistency of temporal resolution across all data sets, hourly mean values were used for further analysis.

A rigorous flagging process was implemented to remove periods with operational issues or anomalies, ensuring data set reliability. To validate the data quality and accuracy of the data set from Dublin site in this study, the reconstructed PM₁ (= NR-PM₁ + eBC) was compared with the collocated SMPS measurements, with the PM₁ mass concentration converted into volume concentration by dividing the mass concentrations of each species by their respective densities.³⁶ Additionally, PM_{2.5} mass concentration measurements from the Rathmines monitoring station located around 3 km from the Dublin site were used as an external comparison data set. The intercomparisons showed high correlation coefficients ranging from 0.83 to 0.97, associated with reasonable slopes (0.84 for PM_{2.5}/PM₁ and 1.15 for PM₁/SMPS, respectively), which validated the Dublin data set (Figure S1b–c). Additionally, the good agreement with PM_{2.5} measurements from Rathmines confirmed that the impacts from nearby local point sources are insignificant at the Dublin sampling site.

2.2. Rolling Positive Matrix Factorization (Rolling PMF). The raw data set collected by Q-ACSM was processed using standard data analysis software (version 1.6.1.1) based on Igor Pro (Wavemetrics Inc.) to obtain the mass concentration of NR-PM₁ species. The concentration and uncertainty matrices of OA were also exported from this software and averaged to hourly resolution, serving as inputs of PMF analysis. To enhance the separation of OA factors originating from different sources, mass profiles of peat (OA from peat burning), wood (OA from wood burning), and coal (OA from coal burning) obtained from burning experiments^{37–39} and the hydrocarbon-like OA (HOA) profile from AMS UMR MS database (<https://cires1.colorado.edu/jimenez-group/AMSsd/>), which was measured in Paris,⁴⁰ were used as reference profiles to constrain the POA factors using ME-2 algorithm under Source Finder pro (SoFi-pro). Taking the dynamic nature of OA factors into consideration, the advanced rolling-PMF analysis, which is able to dynamically adjust the mass profiles over time,⁴¹ was performed, with a rolling time-window of 14 days and shifting step of 1 day, respectively. In addition, segmented PMF analysis and the “limits” a-value approach were applied to further account for potential source variability over time. The constraining strategies and bootstrap resampling applied to the data set in this study are consistent with our previous studies in Dublin.^{39,42} More details are provided in S1.2, and the criteria for PMF run selection are summarized in Table S3.

2.3. Machine Learning Model. **2.3.1. Training Data Set Selection.** A supervised machine learning model was developed (Figure S2) to enhance the differentiation of the OOA into local and transboundary sources in this study. Pollution events dominated by local emissions were selected as the model training data set to establish relationships between OOA from local sources and their corresponding predictor variables. To ensure the reliability of selected local air pollution events, a set of screening criteria (Table S4) was implemented to exclude impacts from regional sources: (1) Previous studies have demonstrated that, in Dublin, the air pollution events caused by local residential heating emissions mainly occur in cold months,⁴² so only data from October to March were selected for model training; (2) Data points when OA mass concentration was lower than 0.5 μg m⁻³ were excluded to avoid high uncertainty associated with lower signal-to-noise ratios; (3) The mass fraction of POA factors to total OA was constrained to be higher than OOA (POA fraction >50%) to ensure the dominance of local emissions; (4) The mass ratio of OA to NO₃ was constrained to be higher than 2 to exclude significant impacts from regional transport, which features high contribution of NO₃;³² and (5) Data points were selected only when WS was below 5 m s⁻¹, as stagnant meteorological conditions favored extreme pollution events. The thresholds of filtering criteria were chosen to balance the selection of representative local events, while ensuring adequate data for reliable model training. Sensitivity tests, as shown in Figure

S3a, confirmed that the model effectively captures the characteristics of local emissions without overly dependent on specific thresholds. As a result, ~5% of the entire data set (2733 out of 56176 data points) was selected for model training and feature selection.

2.3.2. Model Selection, Optimization, and Evaluation. A support vector regression (SVR) model was chosen in this study to predict the OOA from local sources due to its well-documented advantages in handling nonlinear relationships, robustness even with relatively small data sets, and its resistance to overfitting.⁴³ The choice of SVR was further validated through comparisons with other machine learning algorithms (Figure S3b). The selection of model predictors was optimized to ensure the best model performance (Figure S4a,b), with further details provided in the Supplement (S1.3). As summarized in Table S4, primary species (POA, eBC, and Cl) were included as model inputs as they tend to spike concurrently with OOA during pollution events dominated by local emissions.^{29,30} WD was also incorporated in the model, as it provides information about air mass origins, which is crucial for understanding local pollution dynamics. Hour of day was also considered, as local domestic heating emissions follow distinct diurnal patterns,³⁰ offering additional information to differentiate local emissions from transboundary transport. Total OA was chosen to serve as a rough threshold of the model prediction, preventing large deviations from the measured concentrations. In addition, LO-OOA was included as a predictor for MO-OOA due to their common sources during local-emission-dominated periods and the fact that LO-OOA can potentially evolve into MO-OOA through further oxidation.^{6,11,14} Ambient *T* was excluded to avoid potential bias because the training period focused on winter, where *T* is consistently lower compared to summer (Figure S5a). Additionally, local emissions from home heating tend to be higher in winter than during periods in warmer months that have comparable temperatures, while sporadic residential heating during cooler summer periods further complicates the relationship between *T* and local emissions (Figure S5b). RH was excluded since it remains consistently high throughout the year in Ireland (monthly averaged RH > 75%, Figure S5c), providing minimal additional information. WS was not used, because it was already applied in data screening. The selected training data set was randomly split into two parts: 80% for model training and 20% for model testing. The distribution of the number of data points used for model training and testing can be found in Figure S4c. The model was optimized using a grid search, which involves systematically testing a range of hyperparameters and 5-fold cross-validation to ensure reasonable predictions. After evaluation and validation, the model was applied to the entire Dublin data set from 2016 to 2023. Once the local sources-related OOA (LO-OOA_{local} and MO-OOA_{local}) were estimated, the contributions from transboundary transport (LO-OOA_{TBT} and MO-OOA_{TBT}) were extracted accordingly using equations as below:

$$LO-OOA_{TBT} = \begin{cases} LO-OOA_{PMF} - LO-OOA_{local} & \text{if } LO-OOA_{PMF} > LO-OOA_{local} \\ 0 & \text{if } LO-OOA_{PMF} \leq LO-OOA_{local} \end{cases}$$

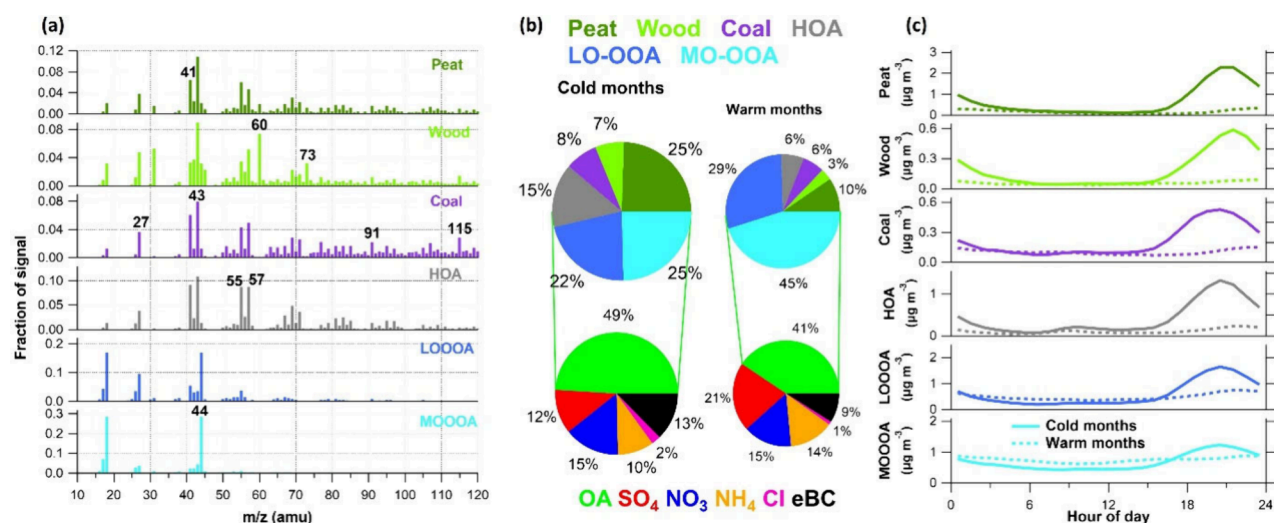


Figure 1. (a) Mass profiles of six OA factors identified by rolling-PMF analysis including Peat, Wood, Coal, HOA, LO-OOA, and MO-OOA, averaged over all segmented analysis periods. (b) The average chemical composition of total PM₁ and OA during cold (October to March) and warm months (April to September) and (c) diurnal variations of OA factors in cold months (solid lines) and warm months (dashed lines) from 2016 to 2023 in Dublin.

$$MO-OOA_{TBT} = \begin{cases} MO-OOA_{PMF} - MO-OOA_{local} & \text{if } MO-OOA_{PMF} > MO-OOA_{local} \\ 0 & \text{if } MO-OOA_{PMF} \leq MO-OOA_{local} \end{cases}$$

The Monte Carlo simulation⁴⁴ was applied to evaluate the model robustness by randomly omitting 20% of the training data and repeating this process 1000 times. In addition, permutation importance analysis was performed for all predictors to identify critical factors that affect the model's accuracy, and the partial dependence plots were used to illustrate the relationship between LO-OOA_{local} and MO-OOA_{local} on their predictors. These analyses help identify how different predictors influence the model outputs and provide additional insights into local sources and the underlying processes driving the local formation of an OOA.

3. RESULTS AND DISCUSSIONS

3.1. OA Identification with Rolling PMF. The six-factor solution was consistently selected as the optimal result for each segmented rolling-PMF analysis over the years. This solution successfully distinguished four POA factors from different fuel types, including Peat, Wood, Coal, and hydrocarbon-like OA (HOA), as well as two OOA factors, i.e., LO-OOA and MO-OOA. Increasing the number of factors did not yield more physically meaningful sources; e.g., as presented in Figure S6, the 8-factor solution resulted in two splitting factors with the r^2 with LO-OOA reaching as high as 1 for profiles and 0.96 for time series.

As depicted in Figure 1a, the mass spectra of HOA exhibited apparently stronger signals at m/z s (mass-to-charge ratios) that are related to hydrocarbon emissions, i.e., C_nH_{2n-1} (m/z 27, 41, 55, 69, etc.) and C_nH_{2n+1} (m/z 43, 57, 71, etc.), and showed overall high correlation with eBC ($r^2=0.74$). It is worth noting that, while HOA is commonly attributed to traffic emissions in many studies,^{15,45} domestic oil burning was identified as a significant source of HOA in Dublin during residential heating hours, characterized by much higher HOA to eBC ratios compared to vehicle emissions.³⁰ The mass

profile of Wood was characterized by significant contributions at m/z 60 (0.07) and m/z 73 (0.03), key tracers of anhydrosugars such as levoglucosan that are produced from the pyrolysis of cellulose during biomass material combustion processes.^{46,47} Coal showed pronounced fractions at m/z s that are tightly correlated with polycyclic aromatic hydrocarbons (PAHs) such as m/z 91 and 115. Peat had a lower but notable fraction at m/z 60 (0.02), consistent with the lower cellulose content in peat.^{29,39} These profile signatures serve as useful markers to distinguish POA factors from different sources. Comparatively, the two OOA factors exhibited much higher signal at m/z 44 (0.19–0.29), which is mostly CO_2^+ serving as a good indicator of atmospheric aging,^{48,49} effectively distinguishing them from POA factors (0.01–0.03), with MO-OOA having highest m/z 44 fraction (0.29). However, both OOA factors are overwhelmingly dominated by m/z s less than 50 (0.71 for LO-OOA and 0.86 for MO-OOA), providing very limited information to further link them to specific sources. Given that segmented PMF analysis was performed for the Dublin data set, we also evaluated whether the mass profiles of OOA factors significantly changed over time. As displayed in Figure S7, despite minor variations in LO-OOA, both OOA factors exhibited highly consistent mass profiles over the years, with MO-OOA remaining particularly stable ($r^2 = 1.0$), indicating the stability of their chemical composition over time.

The pie charts in Figure 1b present the average chemical composition of PM₁ and OA during different seasons, i.e., cold months (October to March) and warm months (April to September) in Dublin from 2016 to 2023. In the cold months, PM₁ was dominated by carbonaceous components (OA and eBC), with an average mass contribution of 62%. In addition, POA contributed a significant fraction to total OA at 53%, with Peat being the most significant contributor (25%), followed by

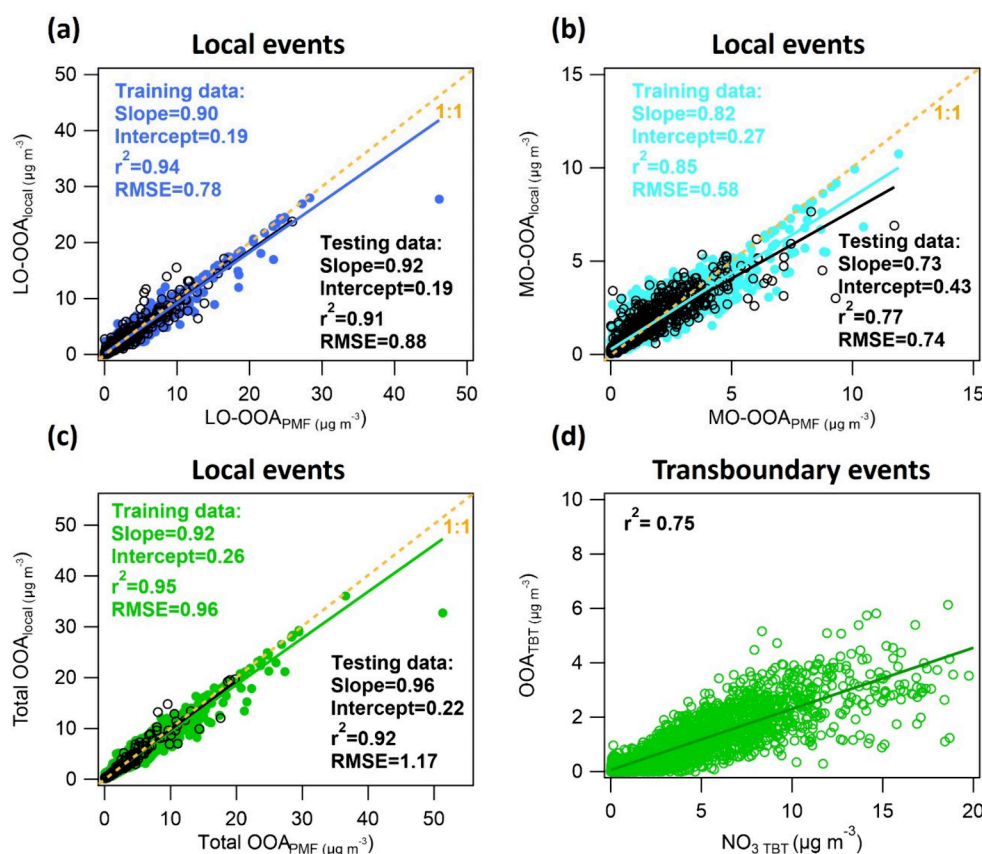


Figure 2. Scatter plots comparing predictions from the model and rolling-PMF analysis for (a) LO-OOA, (b) MO-OOA, and (c) total OOA during selected local events. Panel (d) shows the scatter plot of OOA_{TBT} calculated from the model versus measured NO₃ filtered for transboundary transport (NO₃_{TBT}).

HOA (15%), consistent with previous studies that were conducted in Dublin,³⁰ indicating significant roles of local domestic heating emissions. Meanwhile, OOA factors (LO-OOA and MO-OOA) also shared a significant portion of total OA in this season (47%). It is important to note that the diurnal variations of both LO-OOA and MO-OOA showed high similarities with POA factors (Figure 1c), with significant elevations during heating hours (17:00–24:00), indicating their common sources. Additionally, OOA factors were often observed to concurrently spike with primary species (POA and eBC) during local events,^{29,30} further suggesting the significant contributions from local domestic heating emissions on OOA. This is likely due to the condensation of semivolatile components and the rapid oxidation of primary species.^{29,42}

On the contrary, in warm months, the contributions of secondary inorganic species (SIA, i.e., SO₄ + NO₃ + NH₄) shared equal mass fraction with carbonaceous components. Consistently, OOA contributed more significantly to total OA during warm months (74%), with MO-OOA alone contributing 45% of total OA mass, indicating higher impacts from transboundary transport in this season. In addition, OOA factors showed relatively flat diurnal patterns in warm months, exhibiting typical diurnal variation characteristics of transboundary transport. The results reveal that OOA factors play important roles in causing air pollution events in Dublin, constituting 47–74% of total OA, and are significantly influenced by both local emissions and transboundary transport, with pronounced seasonal variations in composition and sources. However, as noted earlier, traditional PMF

analysis was not able to distinguish between local and transboundary OOA.

3.2. Evaluation of Model Performance. A supervised SVR ML model was developed to enhance differentiation of the OOA into local and transboundary sources. The high similarity in the PM₁ and OA composition for selected training data with that of a representative extreme local pollution event (Figure S8a–b) demonstrates the effectiveness of the screening criteria in accurately extracting local emission-dominated pollution events. Additionally, the diurnal patterns of both LO-OOA and MO-OOA mirrored the variations of primary species (Figure S8c), further validating the fundamental assumption of the ML model that OOA was predominantly originated from local sources during local events.

3.2.1. Overall Model Performance. Figure 2 shows the scatter plots of model predicted local OOA versus PMF-derived OOA during selected pollution events dominated by local domestic heating emissions, including comparisons of LO-OOA, MO-OOA, and total OOA. The prediction of LO-OOA_{local} achieved an r^2 of 0.94 (0.91) for the training (testing) data, which comprises 80% (20%) of the selected data, indicating the ability of the model in effectively capturing the temporary variations of LO-OOA from local sources. The root-mean-square error (RMSE) values, which measure the average magnitude of model prediction errors, were 0.78 $\mu\text{g m}^{-3}$ for training data and 0.88 $\mu\text{g m}^{-3}$ for testing data for LO-OOA, corresponding to 1.7% and 3.4% of the PMF-derived LO-OOA concentrations (0–46 $\mu\text{g m}^{-3}$ for training data and 0–26 $\mu\text{g m}^{-3}$ for testing data, respectively), demonstrating the reliability

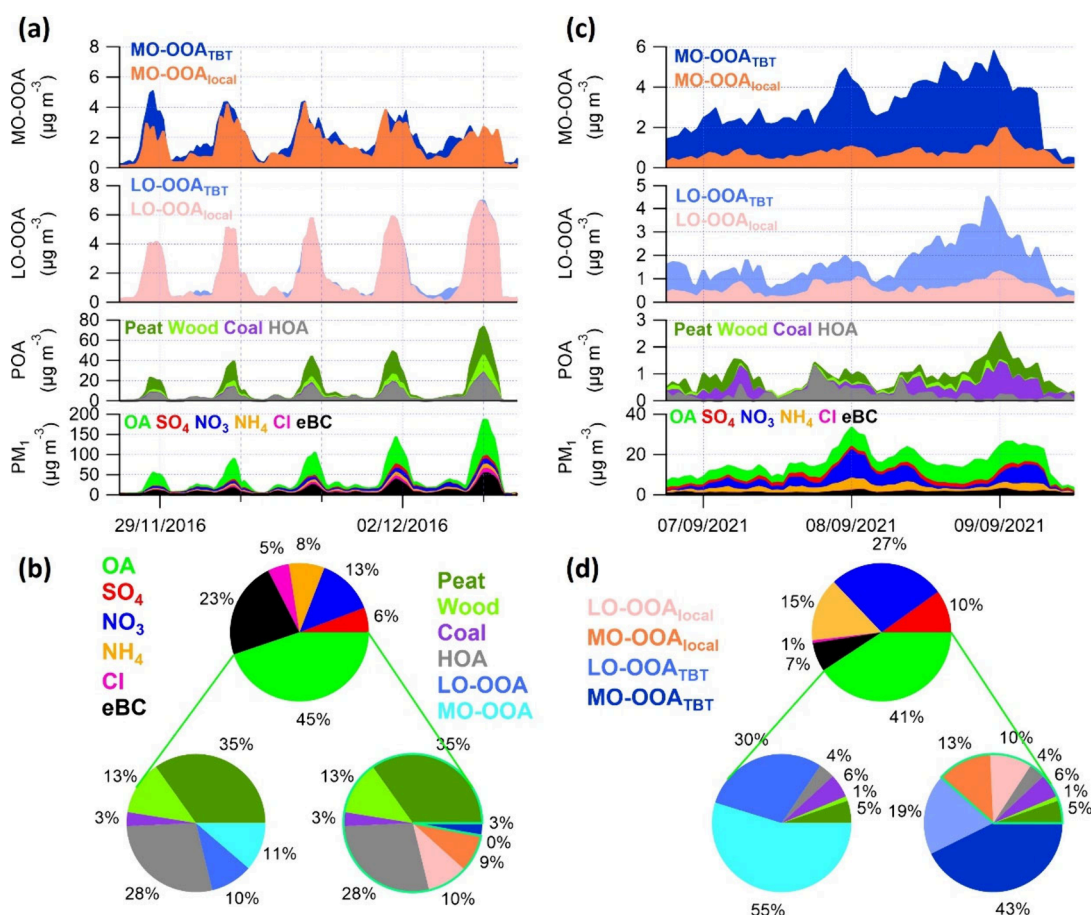


Figure 3. Time series of PM₁ species including OA, SO₄, NO₃, NH₄, Cl, and eBC, and OA factors including Peat, Wood, Coal, HOA, LO-OOA_{local}, LO-OOA_{TBT}, MO-OOA_{local}, and MO-OOA_{TBT} during selected (a) local- and (c) trans-boundary-dominated pollution episodes. Panel (b) presents the average chemical composition for (a) PM₁ (top) and OA composition from rolling PMF analysis (bottom left) and the model (bottom right). Panel (d) provides the same data for episode (c).

of the model predictions. A slope close to 1 (0.90–0.92) further validated the strong agreement between the model predictions and PMF-derived LO-OOA during selected local events. Despite higher uncertainty in the prediction of MO-OOA_{local} compared to that of LO-OOA_{local}, as shown by the scatter plot in Figure 2b, the model achieved an r^2 of 0.85 for training data and 0.77 for testing data. Additionally, the RMSE values for MO-OOA were 0.58 μg m⁻³ for training data and 0.74 μg m⁻³ for testing data, being reasonably low at 4.8% and 8.2% of the respective PMF-derived MO-OOA concentrations (0–12 μg m⁻³ for training data and 0–9 μg m⁻³ for testing data respectively). This suggests that the model prediction of MO-OOA_{local} was reasonable. The higher uncertainty in MO-OOA_{local} may arise from the more complex oxidation processes and diverse origins associated with the highly aged nature of MO-OOA,^{11,14} making it more challenging to establish robust relationships with primary species. Even with higher uncertainty in the MO-OOA prediction, the prediction of total OOA achieved overall strong agreement, with r^2 ranging from 0.92 to 0.95, slopes between 0.92 and 0.96, and low RMSE values of 0.96–1.17 μg m⁻³, corresponding to 1.8–6.5% of the PMF-derived OOA mass concentration. It is worth noting that the ML model predictions tend to be more conservative during periods of extremely high concentrations (OOA > 40 μg m⁻³), leading to slightly underestimated OOA from local sources. This conservative bias can be explained by

the rarity of extreme concentrations, which results in less adequate model training in this high concentration range. The box plots in Figure S9a–b illustrate the model's robustness on a monthly average basis, derived from the Monte Carlo stimulations. The narrow interquartile ranges (25th and 75th percentiles) and short whiskers (10th and 90th percentiles) indicate low variability in the model predictions under various pollution scenarios, especially for LO-OOA_{local}. Additionally, the monthly average standard deviation of LO-OOA_{local} and MO-OOA_{local} ranged from 3 to 6% and 7–13% of their respective concentrations, further confirming the strong robustness of the local OOA predictions throughout the year.

Additionally, the model retrieved OOA originating from transboundary sources was compared to NO₃, a potential tracer of transboundary transport in Ireland.^{31,32} Although its semivolatile nature and contribution of local sources can complicate its temporary variations, NO₃ remains a valuable indicator of transboundary transport, particularly during episodes with high transboundary contributions when the concentration of NO₃ can be equal to or even exceed that of OA. A ratio of NO₃/OA larger than 1.2 was used to filter NO₃ from transboundary transport (NO₃_{TBT}) without significant contributions from local sources. As shown in Figure 2d, the OOA_{TBT} calculated from the model showed a tight correlation with NO₃_{TBT} with r^2 of 0.75, further indicating that the apportionment of OOA_{local} and OOA_{TBT} is robust and

reasonable. This is further supported by their bivariate polar plots. As shown in Figure S10, both LO-OOA_{TBT} and MO-OOA_{TBT} show high concentrations under easterly winds and stagnant conditions, in agreement with previous studies.³¹ In contrast, LO-OOA_{local} and MO-OOA_{local} display distinct local characteristics with no significant dependence on WD, highlighting the effectiveness of the ML model in distinguishing local and transboundary OOA. We also checked instances where OOA_{local} exceeded OOA_{PMF} and found that they mostly (~90%) occurred at very low concentrations (<0.5 $\mu\text{g m}^{-3}$), where PMF uncertainty in separating OOA factors is highest. Given that this low concentration range is not the primary focus of this study, these overfitting occurrences have a negligible impact on the findings.

The permutation importance analysis assessed the relative influences of all predictors on the model predictions of LO-OOA_{local} and MO-OOA_{local}. As depicted in Figure S9c, HOA was found to be the most influential factor on LO-OOA_{local} prediction with a relative importance of 20%, followed by coal (14%), with the relative impacts from all the remaining primary species being significant at 12–13%. While for MO-OOA_{local}, the most influential predictor is LO-OOA (17%), which is expected given their common sources during local events and LO-OOA's potential to form MO-OOA through further oxidation.⁶ The most influential primary species on MO-OOA_{local} are Peat (13%) and Cl (13%). The impacts from WD on both LO-OOA_{local} and MO-OOA_{local} remained insignificant (<1%), further confirming that the Dublin sampling site was not significantly influenced by nearby local point sources.²⁹ It is interesting to note that LO-OOA_{local} prediction is slightly more affected by fossil fuel-related POA (e.g., HOA and Coal) than biomass burning-related POA. This may be because of the lower reactivity of key SOA precursors emitted from fossil fuel combustion (e.g., PAHs) under dark conditions,⁵⁰ leading to slower and less efficient oxidation. Additionally, the majority of fossil fuel OA is typically water-insoluble,^{51,52} limiting further aqueous-phase oxidation. These characteristics may result in fossil fuel emissions being less oxidized and, thus, more closely correlated with LO-OOA. The ubiquitous nonlinear relationships between OOA_{local} and primary species (Figure S11) indicated that other factors, such as oxidants availability, may also significantly affect the formation of OOA, which needs further investigation in the future.

3.2.2. Model Performance During Local and Transboundary Pollution Episodes. To further validate the performance of the model, two distinct air pollution events were selected for detailed verification: one dominated by local emissions and one by transboundary sources, neither of which was included in the training data set. As presented in Figure 3a, five consecutive air pollution episodes occurred from November 29th to December 3rd, 2016, with PM₁ mass concentration peaking from 55 to 186 $\mu\text{g m}^{-3}$. Those pollution events typically started to build up in the late afternoon or early evening (15:00–19:00) when ambient *T* dropped to as low as 0 °C (Figure S12a) and domestic heating started, and gradually dissipated in the early morning hours (00:00 to 05:00). During this period, as shown by the pie charts in Figure 3b, PM₁ was mainly composed of OA (45%) and eBC (23%). OA was dominated by POA factors (79%), with Peat being the most significant contributor (35%), followed by HOA (28%). Comparatively, secondary species, including SIA and OOA, contributed much less significantly; e.g., SIA

accounted for 27% of total PM₁ mass, and OOA shared 21% of total OA mass. Furthermore, as displayed in Figure 3a, both LO-OOA and MO-OOA spiked concurrently with primary species (POA and eBC), indicating a strong connection with local emissions. This is particularly evident for LO-OOA, which showed a high correlation with POA ($r^2 = 0.97$) during this period. The temporary trends and chemical composition clearly highlighted the overwhelmingly dominant role of local emissions.

In agreement with the PMF-derived OOA, which serves as a reference during local-emission-dominated episodes, the model attributed most of the LO-OOA to LO-OOA_{local}, especially during the most polluted heating hours, where the model identified more than 98% of LO-OOA as LO-OOA_{local}. On average, only a negligible portion (4%) of LO-OOA was attributed to transboundary sources, accounting for less than 0.5% of total OA mass throughout this period. Furthermore, LO-OOA_{local} showed almost no correlation with LO-OOA_{TBT} ($r^2 = 0.09$, Figure S13a), confirming that the ML model reliably attributed LO-OOA to its respective local and transboundary sources. For MO-OOA, although a slightly higher fraction (23% on average) was attributed to transboundary sources, the majority of MO-OOA (77%) was linked to local sources, with MO-OOA_{local} typically exceeding 80% of MO-OOA during polluted hours. The lower fraction of MO-OOA_{local} overall could be due to the ubiquitous background and higher model uncertainty caused by its more complex atmospheric processes. Overall, the model showed strong agreement with PMF-derived OOA during air pollution episodes dominated by local domestic heating emissions.

Figure 3c presents an air pollution episode dominated by transboundary transport. During this event, PM₁ started to increase from the evening of September 6th, 2021, from around 6 $\mu\text{g m}^{-3}$ to above 30 $\mu\text{g m}^{-3}$ under easterly winds (Figure S12b), lasting until the noon of September 9th with stagnant meteorological conditions ($WS < 5 \text{ m s}^{-1}$). NO₃ and MO-OOA showed the most significant increases, rising from around 0.5 $\mu\text{g m}^{-3}$ to over 14 $\mu\text{g m}^{-3}$ and 1.5 $\mu\text{g m}^{-3}$ to around 6 $\mu\text{g m}^{-3}$, respectively. On average, PM₁ was predominantly composed of SIA (Figure 3d), with NO₃ being the largest contributor (27%), followed by NH₄ (15%). Although OA remained the most significant component of PM₁ (41%), the contribution from eBC was minor, at 7%. Importantly, OA was overwhelmingly composed of OOA, with MO-OOA alone accounting for 55% of total OA mass, and LO-OOA also contributing significantly at 30%. In contrast, POA factors showed minor contributions, with their average mass fractions ranging from 1 to 6% and mass concentration under 1 $\mu\text{g m}^{-3}$, showing no significant increase during typical heating hours. The significant contributions of NO₃ and MO-OOA, alongside the much smaller fractions of primary species, clearly highlighted the dominant role of transboundary sources. Consistently, the model attributed 65% of LO-OOA and 77% of MO-OOA to transboundary transport, aligning well with the transboundary characteristics of this pollution episode. It is worth noting that, although local emissions contributed much less significantly than transboundary sources during this episode, POA factors still accounted for an average of 16% of total OA mass, peaking at 30% during rush hours when traffic emissions were high. This indicates that local sources may still play a notable role in OOA formation, especially for LO-OOA which is fresher and more strongly influenced by local emissions.⁵³ Both LO-

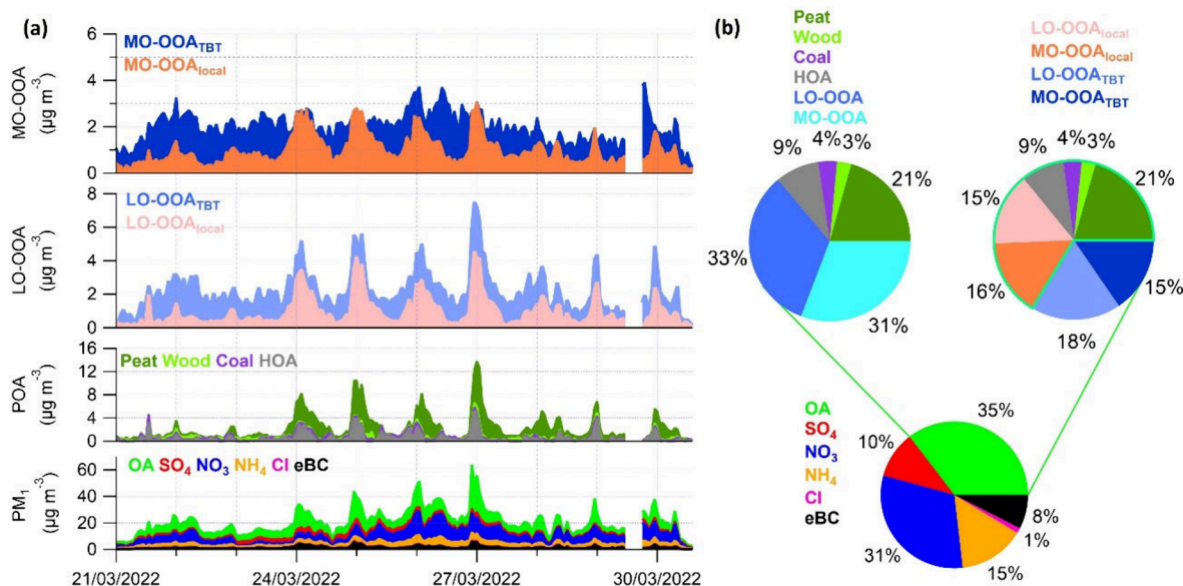


Figure 4. (a) Time series of PM₁ species including OA, SO₄, NO₃, NH₄, Cl, and eBC, and OA factors including Peat, Wood, Coal, HOA, LO-OOA_{local}, LO-OOA_{TBT}, MO-OOA_{local}, and MO-OOA_{TBT} during selected mixed pollution episode in March 2022. Panel (b) shows the average chemical composition of PM₁ (bottom) and OA from rolling PMF analysis (top left) and model (top right) during this episode.

OOA_{local} and MO-OOA_{local} showed tight correlations with POA, with r^2 values of 0.71 and 0.68, respectively, suggesting that the model is able to effectively capture OOA from local sources even under significant impacts of transboundary transport.

3.2.3. Model Application on Mixed Pollution Episode. Mixed air pollution events, where both transboundary transport and local emissions play substantial roles, pose significant challenges for traditional PMF to attribute OOA from different origins. As shown in Figure 4a, an air pollution event started to build up from the evening of March 21st, 2022, with the mass concentrations of NO₃ and OOA significantly increased from $<2 \mu\text{g m}^{-3}$ to $\sim 19 \mu\text{g m}^{-3}$ and $\sim 10 \mu\text{g m}^{-3}$, respectively. Favored by stagnant meteorological conditions (e.g., $WS < 5 \text{ m s}^{-1}$, Figure S14), this pollution episode lasted a few days until the morning of March 30th, 2022. The significant enhancement of NO₃ and OOA, associated with the relatively long duration of this pollution episode, indicate strong impacts from transboundary transport.³¹ Meanwhile, significant contributions from local domestic heating emissions were observed at night, especially from March 24th to 30th when T dropped below $5 \text{ }^\circ\text{C}$. For example, during this period, the mass concentrations of POA and eBC spiked to $8\text{--}14 \mu\text{g m}^{-3}$ and $4\text{--}6 \mu\text{g m}^{-3}$ respectively at night, while their daytime mass concentrations remained below $2 \mu\text{g m}^{-3}$. Overall, as presented in Figure 4b, OA was the most dominant species during this episode, with the average mass fraction of 35%, accompanied by a notable contribution of eBC (8%). Although the contribution from all POA was minor compared to typical local events (36% vs $>50\%$),³⁰ the fraction can reach up to 60% at night, indicating significant influences from local domestic heating. On the other hand, the fractional contribution of NO₃ to total PM₁ was nearly equal to OA at 31%, and the OOA factors dominated total OA with fractions of LO-OOA and MO-OOA being close at 33% and 31%, respectively. The results reveal that both local residential heating emissions and transboundary transport contributed significantly to this episode, but a clear breakdown of their

relative contributions could not be achieved using PMF analysis alone.

The temporal variations of LO-OOA_{local} and MO-OOA_{local} predicted by the ML model showed clear local emission characteristics (Figure 4a), with significant concentration increases only at night, spiking to $4.6 \mu\text{g m}^{-3}$ and $3.0 \mu\text{g m}^{-3}$, respectively. These increases coincided with typical residential heating hours, which were highly consistent with POA and eBC. The correlation coefficients with POA reached 0.96 for LO-OOA_{local} and 0.80 for MO-OOA_{local}, respectively. Comparatively, the r^2 values of PMF-derived LO-OOA and MO-OOA with POA were 0.78 and 0.02, respectively (Figure S15a-b). LO-OOA_{TBT} and NO₃ displayed similar temporal trends, with both showing notable enhancements at night, likely due to their semivolatile properties.^{11,54} Impacts from the shallower nocturnal boundary layer are also expected, and local residential heating emissions may also have played a role in the NO₃ increase. However, it is important to acknowledge that the nighttime enhancement of LO-OOA_{TBT} may be linked to the possible underprediction of LO-OOA_{local}, as the most pronounced LO-OOA_{TBT} increase coincided with the strongest local emissions, despite their overall weak correlation ($r^2 = 0.20$, Figure S13d). MO-OOA_{TBT} showed a strong correlation with NO₃ TBT during daytime ($r^2 = 0.79$, Figure S15c) when residential heating emissions were minimal ($POA < 1.5 \mu\text{g m}^{-3}$), while this correlation weakened at night ($r^2 = 0.57$). The weaker correlation may result from their different volatilities; however, the MO-OOA_{TBT} concentration showed notable decreases during nights when local domestic heating emissions were particularly strong (March 24th to 29th), implying that MO-OOA_{local} might have been slightly overpredicted during these nights. Despite some uncertainties, the results elucidate that the model is capable of providing reasonable quantifications of local and transboundary sources, even during mixed pollution episode.

Although both local and transboundary sources contributed significantly during this mixed pollution episode, their relative importance varied by time of day. During nonheating hours

($\text{POA} < 1.5 \mu\text{g m}^{-3}$), NO_3 contributed slightly more to total PM_1 than OA (33% vs 31%, Figure S16a), and OOA_{TBT} dominated OA with the average fraction at 56%. This qualitatively aligns with the traditional PMF analysis, where OOA constituted 80% of total OA, indicating the dominant role of transboundary transport. In contrast, during heating hours, although NO_3 still contributed significantly at 30% (Figure S16b), the contribution of OA increased to 38%, associated with a notable rise of eBC (6% vs 9%). PMF analysis suggested that OOA still dominated OA at 55%, although POA fractions significantly increased from 20% to 45%. However, the quantitative results from the model elucidated that more than half (61%) of the OOA originated from local emissions. On average, OA from local sources (POA , $\text{LO-OOA}_{\text{local}}$ and $\text{MO-OOA}_{\text{local}}$) accounted for 78% of total OA, highlighting the more critical role of local emissions during heating hours. Over the span of this mixed pollution episode, nearly half of both LO-OOA (45%) and MO-OOA (50%) were emitted from local sources, and the remaining contributions came from transboundary transport (Figure 4b). Over the mixed pollution episode, OA originated from local sources contributed 68% of total OA despite the strong influences from transboundary sources, indicating that local emissions played more critical roles, especially at night. However, since the traditional PMF analysis showed dominance of OOA (64%), the lack of detailed OOA origin information could lead to contrary conclusions about transboundary sources, ultimately resulting in ineffective control strategies.

4. CONCLUSIONS

This study demonstrated the capability and effectiveness of a machine learning model in enhancing the source attribution of OOA for the Dublin data set from 2016 to 2023. While the rolling-PMF analysis revealed that the OOA accounted for a significant fraction of total OA in Dublin (47–74%), it lacked the ability to differentiate the OOA from different sources. In contrast, the machine learning model successfully distinguished OOA contributions from local and transboundary sources, providing quantitative insights into their relative impacts. The model exhibited strong agreements with PMF-derived OOA during local emission-dominated events, with robust predictions reflected by low RMSE values and further supported by Monte Carlo simulations. During events dominated by local emissions, the model successfully attributed most of the LO-OOA (96%) and MO-OOA (77%) to local sources. Similarly, during pollution episodes dominated by transboundary transport, the model effectively attributed the majority of the OOA (66–77%) to transboundary sources. The relative importance analysis indicated that $\text{LO-OOA}_{\text{local}}$ prediction was more affected by fossil fuel emissions, such as HOA (20%) and coal (14%), while $\text{MO-OOA}_{\text{local}}$ was primarily influenced by LO-OOA (17%), implying their potential sources and formation mechanisms. By applying the model to a mixed pollution episode, the findings underscored that despite substantial contributions from transboundary transport during this episode, local emissions from residential heating were more critical sources of OA, with local OA on average accounting for 68% of total OA and reaching 78% during heating hours. This highlights the continued need to reduce local OA emissions and the importance of distinguishing local OOA sources from transboundary transport for effective air pollution control strategies. The successful application of the

machine learning model in this study demonstrated the significant advantages of machine learning in enhancing the OOA source apportionment, with potential for broader applications beyond the OOA. However, this machine learning method also has certain limitations. While it effectively differentiates OOA from local and transboundary sources, incorporating oxidants data (e.g., O_3 and NO_x) could further reduce model uncertainty, especially for MO-OOA. The approach currently relies on observational and statistical methods, but validation against modeling results could further strengthen confidence in its applicability. In addition, although trained on a valuable multiyear data set (2016–2023), regular updates with new ambient data are needed to account for variability in emission sources over time.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsestair.4c00331>.

Details of NR- PM_1 measurements in Dublin (S1.1), details of the rolling PMF analysis (S1.2), ML model evaluation and optimization (S1.3), detailed meteorological conditions (Figures S5, S12, and S14), 8-factor rolling PMF results (Figure S6), comparisons of the training data set with local events (Figure S8), uncertainty analyses for model predictions (Figure S9a–b) and relative importance analysis of predictors (Figure S9c–d), analysis of partial dependences of predictors (Figure S11), correlation plots for OOA factors and other key species (Figure S15), and average OA and PM_1 composition during different hours (Figure S16) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Wei Xu – Center for Excellence in Regional Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China; orcid.org/0000-0002-9590-1906; Email: wuxu@iue.ac.cn

Jurgita Ovadnevaite – School of Natural Sciences, Ryan Institute's Centre for Climate & Air Pollution Studies, University of Galway, Galway H91 CFS0, Ireland; Email: jurgita.ovadnevaite@universityofgalway.ie

Authors

Lu Lei – School of Natural Sciences, Ryan Institute's Centre for Climate & Air Pollution Studies, University of Galway, Galway H91 CFS0, Ireland; orcid.org/0009-0002-1899-9580

Chunshui Lin – State Key Laboratory of Loess and Quaternary Geology and Key Laboratory of Aerosol Chemistry and Physics, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

Baihua Chen – Center for Excellence in Regional Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

Kirsten N. Fossum – School of Natural Sciences, Ryan Institute's Centre for Climate & Air Pollution Studies, University of Galway, Galway H91 CFS0, Ireland

Darius Ceburnis – School of Natural Sciences, Ryan Institute's Centre for Climate & Air Pollution Studies, University of Galway, Galway H91 CFS0, Ireland

Colin O'Dowd – School of Natural Sciences, Ryan Institute's Centre for Climate & Air Pollution Studies, University of Galway, Galway H91 CFS0, Ireland

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acsestair.4c00331>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has received funding from European Union's Horizon Europe Research and Innovation programme under HORIZON-CL5-2022-D1-02 (grant no. 101081430 – PARIS). This publication has also emanated from research jointly funded by Taighde Éireann – Research Ireland under Grant number [22/FFP-A/10611], and by National Key R&D Program of China (2023YFC3705503), Hundred Talents Program of Chinese Academy of Sciences, EPA-Ireland, the Department of Environment, Climate and Communications.

REFERENCES

- (1) Myhre, G.; Myhre, C. L. E.; Samset, B. H.; Storelvmo, S. Aerosols and their Relation to Global Climate and Climate Sensitivity. *Nature Education Knowledge* **2013**, *4* (5), 7.
- (2) Arias, P.; Bellouin, N.; Coppola, E.; Jones, R.; Krinner, G.; Marotzke, J.; Naik, V.; Palmer, M.; Plattner, G.-K.; Rogelj, J. *Climate Change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; technical summary; 2021.
- (3) Pope, C. A.; Dockery, D. W. Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Manage. Assoc.* **2006**, *56* (6), 709–742.
- (4) Shiraiwa, M.; Ueda, K.; Pozzer, A.; Lammel, G.; Kampf, C. J.; Fushimi, A.; Enami, S.; Arangio, A. M.; Frohlich-Nowoisky, J.; Fujitani, Y.; Furuyama, A.; Lakey, P. S. J.; Lelieveld, J.; Lucas, K.; Morino, Y.; Poschl, U.; Takahama, S.; Takami, A.; Tong, H.; Weber, B.; Yoshino, A.; Sato, K. Aerosol health effects from molecular to global scales. *Environ. Sci. Technol.* **2017**, *51* (23), 13545–13567.
- (5) Arfin, T.; Pillai, A. M.; Mathew, N.; Tirpude, A.; Bang, R.; Mondal, P. An overview of atmospheric aerosol and their effects on human health. *Environmental Science and Pollution Research* **2023**, *30* (60), 125347–125369.
- (6) Jimenez, J. L.; Canagaratna, M. R.; Donahue, N. M.; Prevot, A. S. H.; Zhang, Q.; Kroll, J. H.; DeCarlo, P. F.; Allan, J. D.; Coe, H.; Ng, N. L.; Aiken, A. C.; Docherty, K. S.; Ulbrich, I. M.; Grieshop, A. P.; Robinson, A. L.; Duplissy, J.; Smith, J. D.; Wilson, K. R.; Lanz, V. A.; Hueglin, C.; Sun, Y. L.; Tian, J.; Laaksonen, A.; Raatikainen, T.; Rautiainen, J.; Vaattovaara, P.; Ehn, M.; Kulmala, M.; Tomlinson, J. M.; Collins, D. R.; Cubison, M. J.; Dunlea, J.; Huffman, J. A.; Onasch, T. B.; Alfarra, M. R.; Williams, P. I.; Bower, K.; Kondo, Y.; Schneider, J.; Drewnick, F.; Borrmann, S.; Weimer, S.; Demerjian, K.; Salcedo, D.; Cottrell, L.; Griffin, R.; Takami, A.; Miyoshi, T.; Hatakeyama, S.; Shimoono, A.; Sun, J. Y.; Zhang, Y. M.; Dzepina, K.; Kimmel, J. R.; Sueper, D.; Jayne, J. T.; Herndon, S. C.; Trimborn, A. M.; Williams, L. R.; Wood, E. C.; Middlebrook, A. M.; Kolb, C. E.; Baltensperger, U.; Worsnop, D. R. Evolution of organic aerosols in the atmosphere. *Science* **2009**, *326* (5959), 1525–1529.
- (7) Mauderly, J. L.; Chow, J. C. Health effects of organic aerosols. *Inhalation Toxicology* **2008**, *20* (3), 257–88.
- (8) Wu, D.; Zheng, H.; Li, Q.; Jin, L.; Lyu, R.; Ding, X.; Huo, Y.; Zhao, B.; Jiang, J.; Chen, J.; Li, X.; Wang, S. Toxic potency-adjusted control of air pollution for solid fuel combustion. *Nature Energy* **2022**, *7* (2), 194–202.
- (9) Chaturvedi, S.; Kumar, A.; Singh, V.; Chakraborty, B.; Kumar, R.; Min, L. Recent Advancement in Organic Aerosol Understanding: a Review of Their Sources, Formation, and Health Impacts. *Water, Air, & Soil Pollution* **2023**, *234* (12), 750.
- (10) Hallquist, M.; Wenger, J. C.; Baltensperger, U.; Rudich, Y.; Simpson, D.; Claeys, M.; Dommen, J.; Donahue, N. M.; George, C.; Goldstein, A. H.; Hamilton, J. F.; Herrmann, H.; Hoffmann, T.; Iinuma, Y.; Jang, M.; Jenkin, M. E.; Jimenez, J. L.; Kiendler-Scharr, A.; Maenhaut, W.; McFiggans, G.; Mentel, Th. F.; Monod, A.; Prévôt, A. S. H.; Seinfeld, J. H.; Surratt, J. D.; Szmigielski, R.; Wildt, J. The formation, properties and impact of secondary organic aerosol: current and emerging issues. *Atmospheric Chemistry and Physics* **2009**, *9*, 5155–5236.
- (11) Zhang, Q.; Jimenez, J. L.; Canagaratna, M. R.; Ulbrich, I. M.; Ng, N. L.; Worsnop, D. R.; Sun, Y. Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review. *Analytical & Bioanalytical Chemistry* **2011**, *401* (10), 3045–3067.
- (12) Ulbrich, I. M.; Canagaratna, M. R.; Zhang, Q.; Worsnop, D. R.; Jimenez, J. L. Atmospheric Chemistry and Physics Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data. *Atmospheric Chemistry and Physics* **2009**, *9*, 2891–2918.
- (13) Vasilakopoulou, C. N.; Matrali, A.; Skyllakou, K.; Georgopoulou, M.; Aktypis, A.; Florou, K.; Kaltsonoudis, C.; Siouti, E.; Kostenidou, E.; Blaziak, A.; Nenes, A.; Papagiannis, S.; Eleftheriadis, K.; Patoulias, D.; Kioutsioukis, I.; Pandis, S. N. Rapid transformation of wildfire emissions to harmful background aerosol. *npj Climate and Atmospheric Science* **2023**, *6* (1), 218.
- (14) Ng, N. L.; Canagaratna, M. R.; Zhang, Q.; Jimenez, J. L.; Tian, J.; Ulbrich, I. M.; Kroll, J. H.; Docherty, K. S.; Chhabra, P. S.; Bahreini, R.; Murphy, S. M.; Seinfeld, J. H.; Hildebrandt, L.; Donahue, N. M.; DeCarlo, P. F.; Lanz, V. A.; Prévôt, A. S. H.; Dinar, E.; Rudich, Y.; Worsnop, D. R. Organic aerosol components observed in Northern Hemispheric datasets from Aerosol Mass Spectrometry. *Atmospheric Chemistry and Physics* **2010**, *10* (10), 4625–4641.
- (15) Sun, Y. L.; Zhang, Q.; Schwab, J. J.; Yang, T.; Ng, N. L.; Demerjian, K. L. Factor analysis of combined organic and inorganic aerosol mass spectra from high resolution aerosol mass spectrometer measurements. *Atmospheric Chemistry and Physics* **2012**, *12* (18), 8537–8551.
- (16) Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5*, 111–126.
- (17) Zhong, S.; Zhang, K.; Bagheri, M.; Burken, J. G.; Gu, A.; Li, B.; Ma, X.; Marrone, B. L.; Ren, Z. J.; Schrier, J.; Shi, W.; Tan, H.; Wang, T.; Wang, X.; Wong, B. M.; Xiao, X.; Yu, X.; Zhu, J. J.; Zhang, H. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environ. Sci. Technol.* **2021**, *55* (19), 12741–12754.
- (18) Zhu, J. J.; Yang, M.; Ren, Z. J. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environ. Sci. Technol.* **2023**, *57* (46), 17671–17689.
- (19) Mendez, M.; Merayo, M. G.; Nunez, M. Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review* **2023**, *56*, 10031.
- (20) Peng, Z.; Zhang, B.; Wang, D.; Niu, X.; Sun, J.; Xu, H.; Cao, J.; Shen, Z. Application of machine learning in atmospheric pollution research: A state-of-art review. *Sci. Total Environ.* **2024**, *910*, No. 168588.
- (21) Li, H.; Yang, Y.; Wang, H.; Wang, P.; Yue, X.; Liao, H. Projected Aerosol Changes Driven by Emissions and Climate Change Using a Machine Learning Method. *Environ. Sci. Technol.* **2022**, *56* (7), 3884–3893.
- (22) Verma, P.; Verma, R.; Mallet, M.; Sisodiya, S.; Zare, A.; Dwivedi, G.; Ristovski, Z. Assessment of human and meteorological influences on PM10 concentrations: Insights from machine learning algorithms. *Atmospheric Pollution Research* **2024**, *15* (6), No. 102123.
- (23) Peng, X.; Xie, T.-T.; Tang, M.-X.; Cheng, Y.; Peng, Y.; Wei, F.-H.; Cao, L.-M.; Yu, K.; Du, K.; He, L.-Y.; Huang, X.-F. Critical Role of Secondary Organic Aerosol in Urban Atmospheric Visibility

Improvement Identified by Machine Learning. *Environmental Science & Technology Letters* **2023**, *10* (11), 976–982.

(24) Wang, F.; Zhang, F.; Wang, G.; Wang, Z.; Li, M.; Liang, W.; Gao, J.; Wang, W.; Chen, D.; Feng, Y.; Shi, G. Machine learning and theoretical analysis release the non-linear relationship among ozone, secondary organic aerosol and volatile organic compounds. *Journal of Environmental Sciences (China)* **2022**, *114*, 75–84.

(25) Qiao, X.; Li, X.; Yan, C.; Sarnela, N.; Yin, R.; Guo, Y.; Yao, L.; Nie, W.; Huang, D.; Wang, Z.; Bianchi, F.; Liu, Y.; Donahue, N. M.; Kulmala, M.; Jiang, J. Precursor apportionment of atmospheric oxygenated organic molecules using a machine learning method. *Environmental Science: Atmospheres* **2023**, *3* (1), 230–237.

(26) Heikkinen, L.; Äijälä, M.; Daellenbach, K. R.; Chen, G.; Garmash, O.; Aliaga, D.; Graeffe, F.; Rätty, M.; Luoma, K.; Aalto, P.; Kulmala, M.; Petäjä, T.; Worsnop, D.; Ehn, M. Eight years of sub-micrometre organic aerosol composition data from the boreal forest characterized using a machine-learning approach. *Atmospheric Chemistry and Physics* **2021**, *21* (13), 10081–10109.

(27) Pande, P.; Shrivastava, M.; Shilling, J. E.; Zelenyuk, A.; Zhang, Q.; Chen, Q.; Ng, N. L.; Zhang, Y.; Takeuchi, M.; Nah, T.; Rasool, Q. Z.; Zhang, Y.; Zhao, B.; Liu, Y. Novel Application of Machine Learning Techniques for Rapid Source Apportionment of Aerosol Mass Spectrometer Datasets. *ACS Earth and Space Chemistry* **2022**, *6* (4), 932–942.

(28) Goodman, P. G.; Rich, D. Q.; Zeka, A.; Clancy, L.; Dockery, D. W. Effect of Air Pollution Controls on Black Smoke and Sulfur Dioxide Concentrations across Ireland. *J. Air Waste Manage. Assoc.* **2009**, *59* (2), 207–213.

(29) Lin, C.; Huang, R.-J.; Ceburnis, D.; Buckley, P.; Preissler, J.; Wenger, J.; Rinaldi, M.; Facchini, M. C.; O'Dowd, C.; Ovadnevaite, J. Extreme air pollution from residential solid fuel burning. *Nature Sustainability* **2018**, *1* (9), 512–517.

(30) Lin, C.; Ceburnis, D.; Huang, R.-J.; Xu, W.; Spohn, T.; Martin, D.; Buckley, P.; Wenger, J.; Hellebust, S.; Rinaldi, M.; Facchini, M. C.; O'Dowd, C.; Ovadnevaite, J. Wintertime aerosol dominated by solid-fuel-burning emissions across Ireland: insight into the spatial and chemical variation in submicron aerosol. *Atmospheric Chemistry and Physics* **2019**, *19* (22), 14091–14106.

(31) Ovadnevaite, J.; Lin, C.; Rinaldi, M.; Ceburnis, D.; Buckley, P.; Coleman, L.; Facchini, M. C.; Wenger, J.; O'Dowd, C. *Air pollution sources in Ireland*; Environmental Protection Agency: 2021.

(32) Lin, C.; Ceburnis, D.; Huang, R.-J.; Canonaco, F.; Prévôt, A. S. H.; O'Dowd, C.; Ovadnevaite, J. Summertime Aerosol over the West of Ireland Dominated by Secondary Aerosol during Long-Range Transport. *Atmosphere* **2019**, *10* (2), 59.

(33) Ng, N. L.; Herndon, S. C.; Trimborn, A.; Canagaratna, M. R.; Croteau, P. L.; Onasch, T. B.; Sueper, D.; Worsnop, D. R.; Zhang, Q.; Sun, Y. L.; Jayne, J. T. An aerosol chemical speciation monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol. *Aerosol Sci. Technol.* **2011**, *45* (7), 780–794.

(34) Freney, E.; Zhang, Y.; Croteau, P.; Amodeo, T.; Williams, L.; Truong, F.; Petit, J.-E.; Sciare, J.; Sarda-Estève, R.; Bonnaire, N.; Arumae, T.; Aurela, M.; Bougiatioti, A.; Mihalopoulos, N.; Coz, E.; Artinano, B.; Crenn, V.; Elste, T.; Heikkinen, L.; Poulain, L.; Wiedensohler, A.; Herrmann, H.; Priestman, M.; Alastuey, A.; Stavroulas, I.; Tobler, A.; Vasilescu, J.; Zanca, N.; Canagaratna, M.; Carbone, C.; Flentje, H.; Green, D.; Maasikmets, M.; Marmureanu, L.; Minguillon, M. C.; Prevot, A. S. H.; Gros, V.; Jayne, J.; Favez, O. The second ACTRIS inter-comparison (2016) for Aerosol Chemical Speciation Monitors (ACSM): Calibration protocols and instrument performance evaluations. *Aerosol Sci. Technol.* **2019**, *53* (7), 830–842.

(35) Middlebrook, A. M.; Bahreini, R.; Jimenez, J. L.; Canagaratna, M. R. Evaluation of composition-dependent collection efficiencies for the Aerodyne aerosol mass spectrometer using field data. *Aerosol Sci. Technol.* **2012**, *46* (3), 258–271.

(36) Cross, E. S.; Slowik, J. G.; Davidovits, P.; Allan, J. D.; Worsnop, D. R.; Jayne, J. T.; Lewis, D. K.; Canagaratna, M.; Onasch, T. B. Laboratory and Ambient Particle Density Determinations using Light

Scattering in Conjunction with Aerosol Mass Spectrometry. *Aerosol Sci. Technol.* **2007**, *41* (4), 343–359.

(37) Lin, C.; Ceburnis, D.; Hellebust, S.; Buckley, P.; Wenger, J.; Canonaco, F.; Prevot, A. S. H.; Huang, R. J.; O'Dowd, C.; Ovadnevaite, J. Characterization of Primary Organic Aerosol from Domestic Wood, Peat, and Coal Burning in Ireland. *Environ. Sci. Technol.* **2017**, *51* (18), 10624–10632.

(38) Trubetskaya, A.; Lin, C.; Ovadnevaite, J.; Ceburnis, D.; O'Dowd, C.; Leahy, J. J.; Monaghan, R. F. D.; Johnson, R.; Layden, P.; Smith, W. Study of Emissions from Domestic Solid-Fuel Stove Combustion in Ireland. *Energy Fuels* **2021**, *35* (6), 4966–4978.

(39) Lin, C.; Ceburnis, D.; Trubetskaya, A.; Xu, W.; Smith, W.; Hellebust, S.; Wenger, J.; O'Dowd, C.; Ovadnevaite, J. On the use of reference mass spectra for reducing uncertainty in source apportionment of solid-fuel burning in ambient organic aerosol. *Atmospheric Measurement Techniques* **2021**, *14* (10), 6905–6916.

(40) Crippa, M.; DeCarlo, P. F.; Slowik, J. G.; Mohr, C.; Heringa, M. F.; Chirico, R.; Poulain, L.; Freutel, F.; Sciare, J.; Cozic, J.; Di Marco, C. F.; Elsasser, M.; Nicolas, J. B.; Marchand, N.; Abidi, E.; Wiedensohler, A.; Drewnick, F.; Schneider, J.; Borrmann, S.; Nemitz, E.; Zimmermann, R.; Jaffrezo, J. L.; Prévôt, A. S. H.; Baltensperger, U. Wintertime aerosol chemical composition and source apportionment of the organic fraction in the metropolitan area of Paris. *Atmospheric Chemistry and Physics* **2013**, *13* (2), 961–981.

(41) Canonaco, F.; Tobler, A.; Chen, G.; Sosedova, Y.; Slowik, J. G.; Bozzetti, C.; Daellenbach, K. R.; El Haddad, I.; Crippa, M.; Huang, R.-J.; Furger, M.; Baltensperger, U.; Prévôt, A. S. H. A new method for long-term source apportionment with time-dependent factor profiles and uncertainty assessment using SoFi Pro: application to 1 year of organic aerosol data. *Atmospheric Measurement Techniques* **2021**, *14* (2), 923–943.

(42) Lin, C.; Ceburnis, D.; O'Dowd, C.; Ovadnevaite, J. Seasonality of Aerosol Sources Calls for Distinct Air Quality Mitigation Strategies. *Toxics* **2022**, *10* (3), 121.

(43) Basak, D.; Pal, S.; Patranabis, D. C. Support Vector Regression. *Neural Information Processing-Letters and Reviews* **2007**, *11*, 203–24.

(44) Janssen, H. Monte-Carlo based uncertainty analysis: Sampling efficiency and sampling convergence. *Reliability Engineering & System Safety* **2013**, *109*, 123–132.

(45) Canagaratna, M. R.; Jayne, J. T.; Ghertner, D. A.; Herndon, S.; Shi, Q.; Jimenez, J. L.; Silva, P. J.; Williams, P.; Lanni, T.; Drewnick, F.; Demerjian, K. L.; Kolb, C. E.; Worsnop, D. R. Chase Studies of Particulate Emissions from in-use New York City Vehicles. *Aerosol Sci. Technol.* **2004**, *38* (6), 555–573.

(46) Simoneit, B. R. T.; Schauer, J. J.; Nolte, C. G.; Oros, D. R.; Elias, V. O.; Fraser, M. P.; Rogge, W. F.; Cass, G. R. Levoglucosan, a tracer for cellulose in biomass burning and atmospheric particles. *Atmos. Environ.* **1999**, *33* (2), 173–182.

(47) Bhattarai, H.; Saikawa, E.; Wan, X.; Zhu, H.; Ram, K.; Gao, S.; Kang, S.; Zhang, Q.; Zhang, Y.; Wu, G.; Wang, X.; Kawamura, K.; Fu, P.; Cong, Z. Levoglucosan as a tracer of biomass burning: Recent progress and perspectives. *Atmospheric Research* **2019**, *220*, 20–33.

(48) Aiken, A.; Decarlo, P. F.; Kroll, J. H.; Douglas, W.; Sun, Y.; et al. O/C and OM/OC Ratios of Primary, Secondary, and Ambient Organic Aerosols with High-Resolution Time-of-Flight Aerosol Mass Spectrometry. *Environ. Sci. Technol.* **2008**, *42* (12), 4478–4485.

(49) Duplissy, J.; DeCarlo, P. F.; Dommen, J.; Alfarra, M. R.; Metzger, A.; Barmapadimos, I.; Prevot, A. S. H.; Weingartner, E.; Tritscher, T.; Gysel, M.; Aiken, A. C.; Jimenez, J. L.; Canagaratna, M. R.; Worsnop, D. R.; Collins, D. R.; Tomlinson, J.; Baltensperger, U. Relating hygroscopicity and composition of organic aerosol particulate matter. *Atmospheric Chemistry and Physics* **2011**, *11* (3), 1155–1165.

(50) Samy, S.; Zielinska, B. Secondary organic aerosol production from modern diesel engine emissions. *Atmospheric Chemistry and Physics* **2010**, *10* (2), 609–625.

(51) Miyazaki, Y.; Kondo, Y.; Takegawa, N.; Komazaki, Y.; Fukuda, M.; Kawamura, K.; Mochida, M.; Okuzawa, K.; Weber, R. J. Time-resolved measurements of water-soluble organic carbon in Tokyo.

Journal of Geophysical Research: Atmospheres **2006**, *111* (D23), D23206.

(52) Wozniak, A. S.; Bauer, J. E.; Dickhut, R. M.; Xu, L.; McNichol, A. P. Isotopic characterization of aerosol organic carbon components over the eastern United States. *Journal of Geophysical Research: Atmospheres* **2012**, *117* (D13), D13303.

(53) Zhou, W.; Wang, Q.; Zhao, X.; Xu, W.; Chen, C.; Du, W.; Zhao, J.; Canonaco, F.; Prévôt, A. S. H.; Fu, P.; Wang, Z.; Worsnop, D. R.; Sun, Y. Characterization and source apportionment of organic aerosol at 260 m on a meteorological tower in Beijing, China. *Atmospheric Chemistry and Physics* **2018**, *18* (6), 3951–3968.

(54) Zhou, W.; Xu, W.; Kim, H.; Zhang, Q.; Fu, P.; Worsnop, D. R.; Sun, Y. A review of aerosol chemistry in Asia: insights from aerosol mass spectrometer measurements. *Environmental Science: Processes & Impacts* **2020**, *22* (8), 1616–1653.



CAS BIOFINDER DISCOVERY PLATFORM™

CAS BIOFINDER HELPS YOU FIND YOUR NEXT BREAKTHROUGH FASTER

Navigate pathways, targets, and
diseases with precision

Explore CAS BioFinder

