Environmental Science & Technology

pubs.acs.org/est

Article

Investigating the Sources of Urban Air Pollution Using Low-Cost Air Quality Sensors at an Urban Atlanta Site

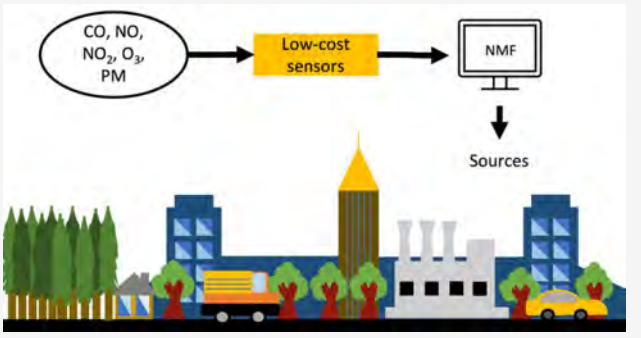
Laura Hyesung Yang, David H. Hagan, Jean C. Rivera-Rios, Makoto M. Kelp, Eben S. Cross, Yuyang Peng, Jennifer Kaiser, Leah R. Williams, Philip L. Croteau, John T. Jayne, and Nga Lee Ng*

Cite This: https://doi.org/10.1021/acs.est.1c07005

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Advances in low-cost sensors (LCS) for monitoring air quality have opened new opportunities to characterize air quality in finer spatial and temporal resolutions. In this study, we deployed LCS that measure both gas (CO, NO, NO_2, and O_3) and particle concentrations and co-located research-grade instruments in Atlanta, GA, to investigate the capability of LCS in resolving air pollutant sources using non-negative matrix factorization (NMF) in a moderately polluted urban area. We provide a comparison of applying the NMF technique to both normalized and non-normalized data sets. We identify four factors with different temporal trends and properties for both normalized and non-normalized data sets. Both normalized and non-normalized LCS data sets can resolve primary organic aerosol (POA) factors identified from research-grade instruments. However, applying normalization provides factors with more diverse compositions and can resolve secondary organic aerosol (SOA). Results from this study demonstrate that LCS not only can be used to provide basic mass concentration information but also can be used for in-depth source apportionment studies even in an urban setting with complex pollution mixtures and relatively low aerosol loadings.



KEYWORDS: *submicron nonrefractory aerosol, aerosol chemical speciation monitor (ACSM), positive matrix factorization (PMF), non-negative matrix factorization (NMF), dimensionality reduction*

INTRODUCTION

The development of low-cost sensors (LCS) to measure air pollutants brings unprecedented opportunities to study air quality, as their lower cost and improved portability allow for a higher spatiotemporal resolution when compared to that of conventional research-grade instruments.[1−3] Recently, LCS have been used to monitor air quality at varying temporal (hourly, daily, and weekly) and spatial (local and regional) scales,[4−7] and there have been active efforts to improve accuracy[8−19] and test the reliability of LCS data.[5,20−22]

Understanding the sources of air pollutants along with their concentrations is useful for abatement and regulatory purposes. Among various air pollutants, understanding the sources of ambient fine particulate matter ($PM_{2.5}$, particles with diameter $\leq$2.5 μm) is of particular concern as its chemical, micro-physical, and optical properties determine the effect on climate,[23] human heath,[24] visibility,[25] and the ecosystem.[26,27] In the last two decades, many studies have investigated the sources and composition of submicron nonrefractory aerosol ($NR\text{-}PM_1$) using the positive matrix factorization (PMF) technique[28] on data obtained from research-grade instruments such as the aerosol mass spectrometer (AMS)[29−40] and the aerosol chemical speciation monitor (ACSM).[41−47] The data from these instruments, combined with factor analysis methods, can deconvolve organic aerosol (OA) (that often dominates fine PM mass) into various OA subtypes, which correlate with different sources and/or atmospheric processing histories. Hydrocarbon-like OA (HOA) and oxygenated OA (OOA) have been resolved as surrogates for primary OA (POA) from direct emission and secondary OA (SOA) from oxidative chemistry, respectively. OOA can be further divided into less-oxidized oxygenated OA (LO-OOA) and more-oxidized oxygenated OA (MO-OOA).[36−38] Other factors such as biomass burning OA (BBOA), cooking OA (COA), and isoprene-derived OA (isoprene-OA) have also been resolved.

Source apportionment analysis using LCS data has been limited. A few studies have attempted to retrieve the sources of particles using LCS data by finding the ratios of different sizes of PM in Kenya,[48] applying non-negative matrix factorization (NMF) to gaseous and particle data in Delhi, India,[49] and

Special Issue: Urban Air Pollution and Human Health

Received: October 15, 2021
Revised: March 15, 2022
Accepted: March 16, 2022

using k-means clustering techniques on the particle size profile at an urban background site in the U.K.[50]

In the computational field, a variety of data scaling, standardization, and transformation (hereafter they are collectively referred to as normalization) techniques are applied to data set.[51] Scaling is applied when input variables have a highly variable data range, and transformation, such as log transformation, is applied to allow skewed data to exhibit a normal distribution.[52] However, such analysis has not been conducted for aerosol source apportionment for LCS.[49]

In this study, we deployed LCS (QuantAQ/Aerodyne Research v200 (hereafter referred to as QuantAQ/ARI v200) and QuantAQ MODULAIR-PM (price $7,995 and $1,295, respectively), capable of measuring both gaseous and particle-phase species. The LCS were colocated with research-grade instrumentation at an urban site, the Georgia Institute of Technology campus in Atlanta, GA, in winter 2020. NMF is applied to the normalized and non-normalized gaseous and aerosol data retrieved from the LCS, and PMF is applied to the ACSM data for source apportionment of aerosol. The method of applying the NMF technique to LCS data follows a similar approach as a previous study that took place in Delhi, India.[49] Here, we test the capability of conducting source apportionment using LCS even in a moderately polluted urban area and investigate the effect of normalization prior to applying the NMF technique. The factors resolved from NMF analysis of both normalized and non-normalized LCS data are compared to the data from research-grade instruments to validate whether LCS can resolve different pollution sources. We also provide a comparison of the results from normalized and non-normalized data. This is the first study to apply normalization before conducting source apportionment analysis on the LCS data. This study provides insights into the advantages and disadvantages of applying normalization prior to NMF analysis and illustrates how LCS data combined with NMF analysis allow for evaluation of pollution sources in a moderately polluted urban area.

MATERIALS AND METHODS

Site Description. Measurements were taken from November 5, 2020, to January 1, 2021. The LCS and the research-grade instruments were simultaneously deployed on the top floor of the Ford Environmental Science & Technology (ES&T) building (height, 30−40 m) on the Georgia Institute of Technology campus. The LCS were placed outdoors on the roof close to the sampling inlet of the research-grade instruments. The research-grade instruments were placed inside the laboratory. The site is characterized as an urban site with strong biogenic influences. We have conducted multiple studies at this site to evaluate the air quality of Atlanta, GA.[35,53,54]

Instrumentation. The research-grade instruments included a time-of-flight ACSM (Aerodyne Research, Inc.)[46] which measured NR-PM_1 composition (organics, sulfate, nitrate, ammonium, chloride) and mass concentrations, and a scanning mobility particle sizer (SMPS, TSI Inc.) consisting of a differential mobility analyzer (DMA) (TSI 3040) and condensation particle counter (CPC) (TSI 3775) which measured the particle size distribution, number, and volume concentrations for mobility diameters between 17.2 nm and 1.0 μm. The PM_1 volume concentration measured by the SMPS was converted to mass concentration by using a weighted average of the density of the individual species measured by the ACSM.[54] The ACSM and SMPS shared the same 1/2 in. stainless steel sampling line (10 m) going from inside the building, where the instruments were located, to the outside where ambient air was sampled. A $PM_{2.5}$ cyclone was used at the end of the sampling line to remove large particles, dust, and debris. The incoming air (4 slm) was pulled using a pump and dried with a Nafion dryer after which the flow was split evenly (2 slm) between the SMPS and the ACSM, employing isokinetic inlets for each instrument. Composition-dependent collection efficiency (CDCE) was applied to the ACSM data to calculate mass concentrations.[55] A chemiluminescence $NO-NO_2-NO_x$ analyzer (model 42C, Thermo Scientific) and an UV photometric O_3 analyzer (model 49C, Thermo Scientific) were used to measure NO_x and O_3 concentrations, respectively. Research-grade NO data were unavailable due to a malfunctioning component inside the instrument. The research-grade NO_2 measurements had an accuracy of ±13% as determined by calibrations using an Airgas certified gas standard with an uncertainty of ±10%. All the research-grade instruments provided 1-min resolution data except for the ACSM (40-s resolution) and SMPS (6-min resolution). Therefore, all data used for this study were averaged to 6 min for the analysis.

Outputs from two types of LCS were used in this study (1-min resolution): (1) electrochemical sensor measurements of CO, NO, NO_2, and O_3 captured with the QuantAQ/ARI v200 sensor system and (2) optical particle counter (OPC-N3) data from the QuantAQ MODULAIR-PM. The QuantAQ/ARI v200 sensor system is composed of five electrochemical sensors (Alphasense CO-B4, NO-B4, NO2-B43F, OX-B431, and CO2 IRC-A1). The O_3 value is obtained by subtracting the NO_2 sensor signal measured by NO2-B43F from the Ox sensor signal measured by the OX-B431 sensor. Previous work provides the calibration of these electrochemical sensors in detail.[13] For the NMF analysis, raw sensor data for the gaseous species (CO, NO, NO_2, and O_3) in the units of millivolts (mV) and raw particle bin data (bin 0, 0.35−0.46 μm; bin 1, 0.46−0.66 μm; and bin 2, 0.66−1.0 μm) in the units of number of particles/cm^3 are used. The raw data of the gaseous species are obtained by subtracting the signal of the auxiliary electrode (AE) from the working electrode (WE) of each electrochemical sensor, a difference that is proportional to the volume mixing ratio of the target analyte.[56] Both types of LCS provide relative humidity (RH) and temperature measurements (Bosch BME280). The average RH is 49 (±18) % and the average temperature is 16 (±7) °C during the study. A previous study has illustrated how the aerosol particle size distribution measured by MODULAIR-PM is minimally affected by RH over the range measured here (Supporting Information section 1).[49] However, electrochemical sensors are susceptible to rapid changes in temperature and RH,[13,15,16,57] and the signals are affected by cross-interferences from other chemical species.[3,14,58,59] Open-source software Python packages are used for data analysis, including seaborn,[60] pandas,[61,62] numpy,[63] and scikit-learn.[64]

To resolve different pollutant factors from the LCS data, NMF is applied.[65,66] NMF has been applied in previous air pollution studies.[49,67−69] To resolve OA factors from the ACSM, PMF[28,70] is applied to the organic portion of the mass spectrum measured by the ACSM. The organic mass spectral and error matrices are exported from the Tofware v3.2.40209 ACSM data analysis software. PMF is performed following the guidelines provided by Zhang et al.[34] and using the PMF

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

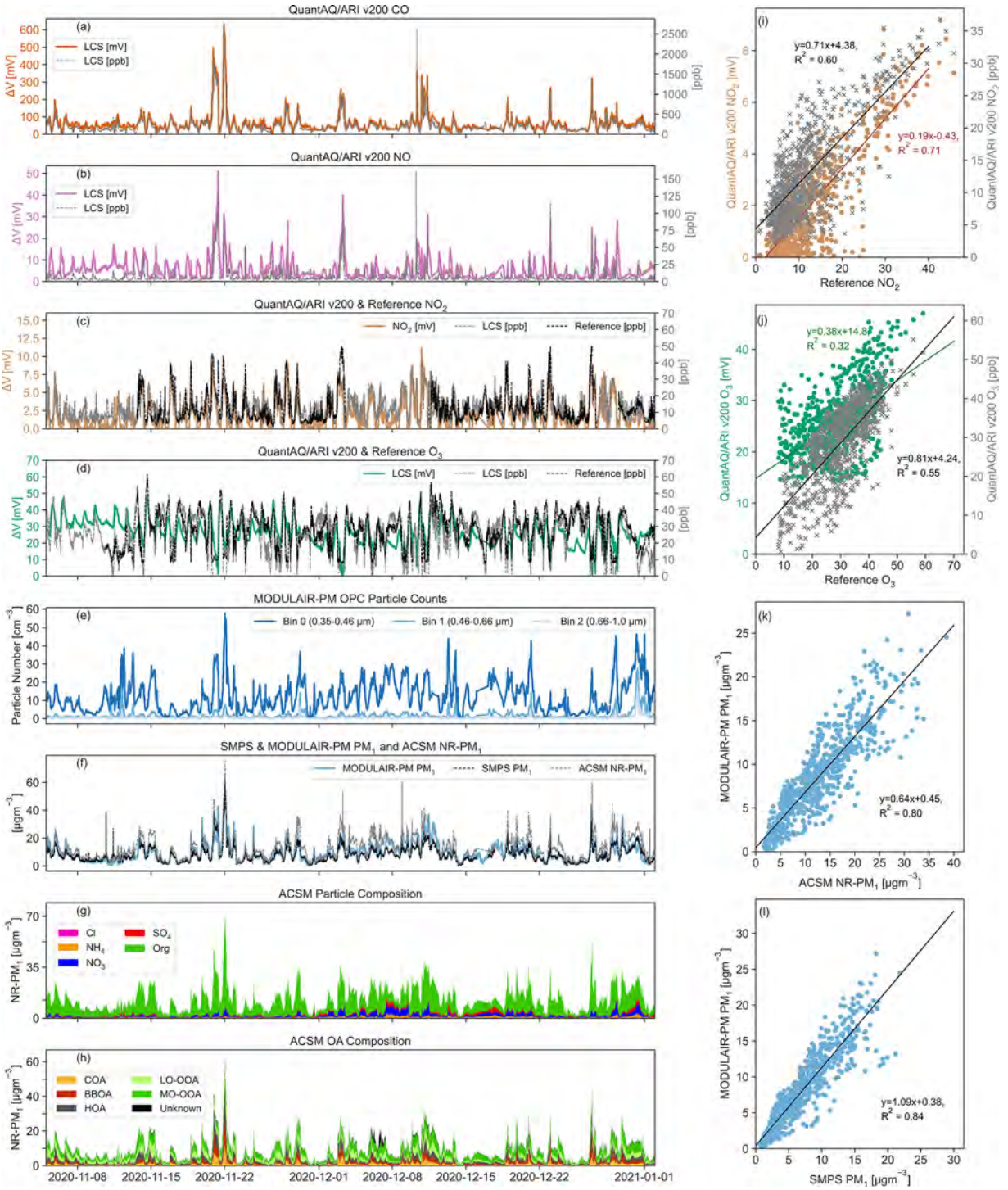


Figure 1. On the left, time series for (a) CO, (b) NO from QuantAQ/ARI v200, (c) NO_2 from QuantAQ/ARI v200 and Chemiluminescence NO−NO_2−NO_x analyzer, (d) O_3 from QuantAQ/ARI v200 and UV photometric O_3 analyzer, (e) OPC bin 0−2 from MODULAIR-PM, (f) integrated PM_1 from MODULAIR-PM, ACSM, and SMPS, (g) Org, SO_4, NO_3, NH_4, and Chl from ACSM, and (h) COA, BBOA, HOA, LO-OOA, and MO-OOA resolved from ACSM PMF analysis. On the right, scatterplots of (i) QuantAQ/ARI v200 NO_2 vs reference NO_2, (j) QuantAQ/ARI v200 O_3 vs reference O_3, (k) MODULAIR-PM PM_1 vs ACSM NR-PM_1 mass concentrations, and (l) MODULAIR-PM PM_1 vs SMPS PM_1 mass concentrations.

Evaluation Tool v 3.06A and PMF2 solver. We downweight the m/z signals with signal-to-noise ratios below 2 by a factor of 2. Errors related to m/z 44 (m/z 17, m/z 18, and m/z 28) are downweighted to avoid repeated information.[29] The optimal solution is chosen based on (1) time series and diurnals of the factors, (2) similarity to previously published factors, (3) comparison to external instruments, and 4) other considerations such as the slope of Q/Q_{exp} and the time series of residuals. Rotational ambiguity is investigated by varying the FPEAK parameter and the robustness of the solutions are examined by running PMF with different starting points (SEED parameter).[34] FPEAK = 0 results in the lowest Q/Q_{exp}

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

and the most representative solution (Supporting Information section 2).

Normalization. The NMF technique has an advantage over PMF for the LCS data as it does not require an error matrix which is not provided by the LCS data. Moreover, NMF is accessible to the public as it is open source, whereas PMF requires licensed software to run the analysis. The advantage of the PMF technique is that it scales each matrix element with its individual error whereas the NMF technique weights each matrix element equally. To address the latter shortcoming, we applied normalization. To compare the effect of normalizing the data prior to employing the NMF technique, NMF is applied to both normalized and non-normalized data. For non-normalized data, the gaseous (CO, NO, NO_2, and O_3) and particle (bins 0−2) raw data are used as input data for the NMF analysis. For the normalized data, first the natural log transformation is applied to each species as they exhibit log-normal distributions (except for O_3 which is already normally distributed). Then, min-max normalization from the scikit-learn Python library is applied to each individual species to make the concentration of all the species range from 0 to 1. Up to these two steps, the purpose is to make all of the data resemble a normal distribution and to ensure that particle data and gaseous data are equally weighted. Next, a weighting factor is applied to individual species according to their original mean values to recreate the variation among gaseous species and among particle species. To determine the weight, the lowest mean values for a gas species (i.e., NO_2) and a particle bin (i.e., bin 2) are used as the baselines. Compared to NO_2, the mean value of CO is 32 times higher, that of O_3 which is 14 times higher, and that of NO which is 3 times higher. Similarly, the mean value of bin 0 is 25 times higher than bin 2's mean value, and that of bin 1 is 16 times higher than bin 2's mean value. Hence, a weighting factor of 32 is applied to the normalized ln(CO), 14 to the normalized O_3, and 3 to the normalized ln(NO). A similar procedure takes place for bin 0 and bin 1 of the particle data. The details of the data distribution and protocol for normalization can be found in the Supporting Information section 3. By normalizing data this way, we (1) reduce the differences in scale between the gaseous and particle species and (2) maintain the variation of magnitude among gaseous species and among particle species.

RESULTS AND DISCUSSION

Comparison of Species Measured by LCS and Research-Grade Instruments. Figure 1 depicts the measured species from LCS and gaseous research-grade instruments. The average NR-PM_1 mass concentration measured by the ACSM is 13 $\mu g\ m^{-3}$. Organics dominate the composition of NR-PM_1 accounting for 77% (10 $\mu g\ m^{-3}$) of the total NR-PM_1. The second most dominant species is nitrate (12%, 1.5 $\mu g\ m^{-3}$), followed by ammonium (6%, 0.72 $\mu g\ m^{-3}$) and sulfate (5%, 0.68 $\mu g\ m^{-3}$). It is noted that November 21−22, 2020 was the weekend before Thanksgiving, which accounts for the spikes in all air pollutants. In particular, the POA tripled during this period compared to the whole campaign period. A previous study in Atlanta reported that long-distance travel on Thanksgiving weekend increases by 54% compared to the annual weekend average.[71] Hence, the traffic during this period can partially explain this spike. Furthermore, this spike is characterized by a period of low wind speeds likely resulting in low boundary layer heights and accumulation of pollutants. There are significant enhancements in aerosol mass concentrations during this time. Other periods of low wind speeds lead to smaller yet significant enhancements in aerosol mass concentrations at nighttime. The diurnal patterns of species measured by the LCS and ACSM can be found in the Supporting Information section 4.

Cross comparisons of species measured by the LCS and the research-grade instruments are also shown in Figure 1. Figure 1 shows the general performance of the LCS used in the current study (Supporting Information section 5), before and after converting the data from the raw signal in mV to concentrations in ppb. This conversion uses a machine learning model developed for each individual sensor.[9,13] It is noted that for the NMF analysis on LCS, only the raw data is used as such data is readily available to users without access to the complex machine learning model,[9,13] which is required for unit conversion. For gaseous species, NO_2 (panel i) and O_3 (panel j) are compared to the research-grade instruments. Both NO_2 and O_3 in raw data format (units of mV) exhibit coefficient of determination (R^2) greater than 0.3 with NO_2 showing a high correlation of 0.71. For O_3, when the unit is converted to ppb, R^2 increases to 0.55. To evaluate the performance of MODULAIR-PM in measuring particle mass concentration, the number concentration is converted to the mass concentration and compared to the ACSM NR-PM_1 (panel k) and the SMPS PM_1 (panel l) mass concentrations. The final MODULAIR-PM PM_1 mass concentrations in the unit of $\mu g/m^3$ calculated using OPC and nephelometer data are comparable to those of ACSM and SMPS PM_1 mass concentrations as R^2 are both $\geq$0.8 with slope >0.6. It should be noted that the nephelometer data are only used for this performance evaluation and are not used in the NMF analysis. Hagan and Kroll provide operating principles and potential limitations of the nephelometer and OPC.[72]

PMF analysis of the ACSM OA data resolves six factors (five factors shown in Figure 1h and all six discussed in the Supporting Information section 2). Five of the factors, HOA, COA, BBOA, LO-OOA, and MO-OOA, are consistent with results from our prior studies at the same site and in the southeastern U.S. in general.[35,53,54,64,73] POA (sum of HOA, COA, and BBOA) and OOA (sum of LO-OOA and MO-OOA) account for 47% and 53% of total OA, respectively. The high level of POA in the study period (winter) arises from reduced photochemical activity and biogenic emissions in the colder months, as we have observed in prior studies in different seasons in the southeastern U.S. The spikes in OA concentration observed during the night of November 21−22 are mostly composed of POA (up to 85% mass fraction). In terms of OOA, we have shown in previous studies that LO-OOA largely arises from oxidation of monoterpenes.[74] In addition to the above factors, we also resolve a factor with prominent signals at m/z 56 and 86 that is only present in significant amounts (16% of total organic signal) from December 6−7 (designated as the Unknown (Unk) factor) (Supporting Information section 2). This factor could arise from some local sources that are only prevalent during those two days.

Factors Resolved from NMF Analysis: Comparison between without and with Normalization Techniques. NMF is applied to four gaseous species (CO, O_3, NO, and NO_2) and three particle bins (0−2) measured by the LCS. To test the reproducibility and estimate the uncertainty of the NMF result, bootstrap analysis[75] is employed. The 4-factor solution is chosen as the best result for both the without and

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

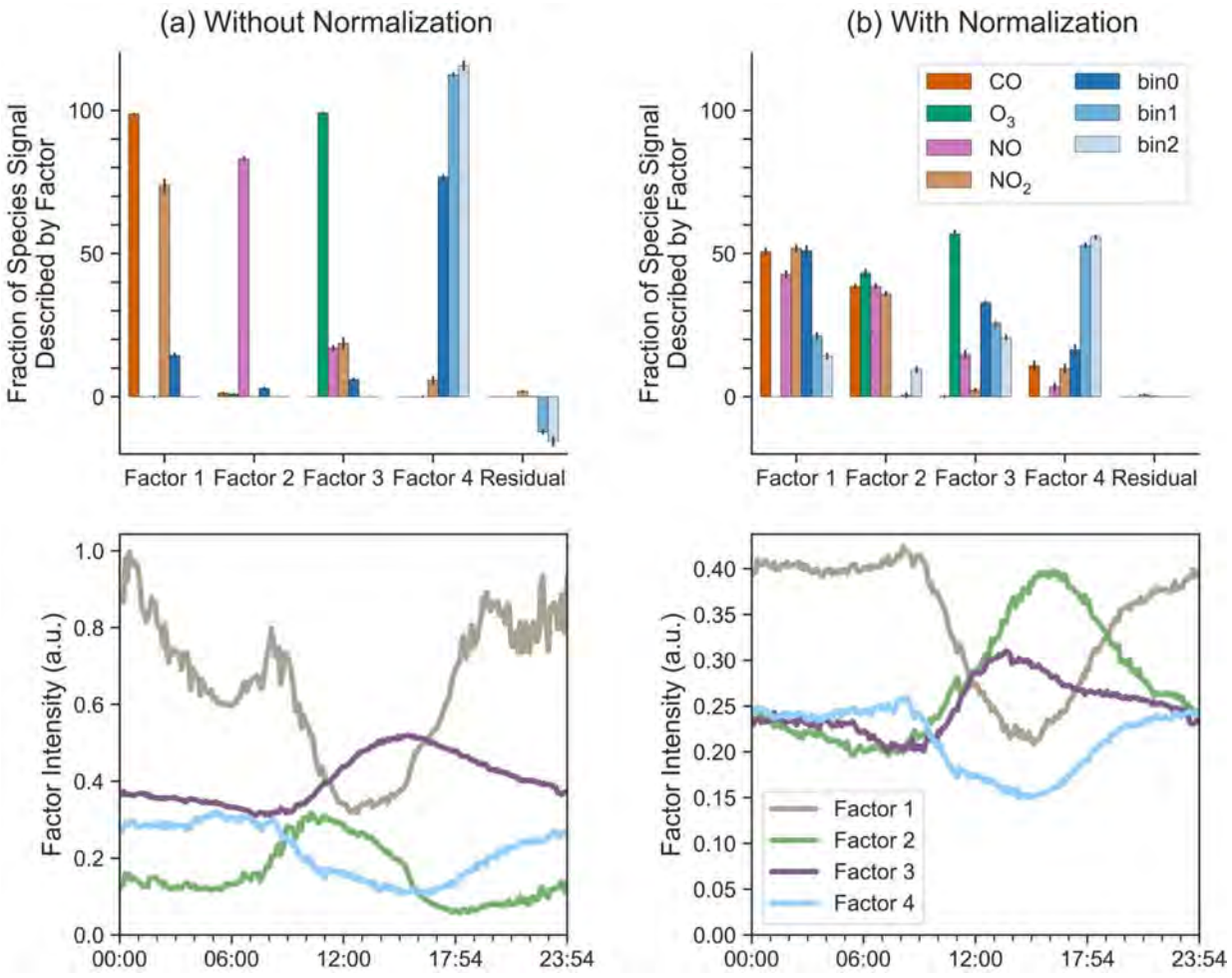


Figure 2. NMF bootstrap results (top row) and diurnal patterns (bottom row) of resolved factors in two different cases: (a) without normalization and (b) with normalization. The fraction of seven species (%) measured by the LCS is distributed across four factors. The residual shows whether the model overestimates (<0) or underestimates (>0). The *x*-axis of the diurnal pattern shows the time from 00:00 AM to 23:54 PM local time.

with normalization cases and is shown in Figure 2. Up to five factors are tested, and the results of 2-, 3-, and 5-factor solutions are presented in the Supporting Information section 6. The selection of the 4-factor solution is based on the cross-validation result[49,76] and the physical meaningfulness in terms of the interpretation of the factors.[28,29,77] The bootstrap analysis and cross-validation approach used in this study are described in Hagan et al.,[49] and the cross-validation result is detailed in the Supporting Information section 7. The result illustrates that when using four factors, the test data show the lowest mean squared error, and this is the primary reason why four factors are chosen as the optimal number of factors for this study. It is noted that the mean squared error for other numbers of factors does not deviate significantly from that of the four factors, so they are all feasible solutions. However, in case the user does not have a reference instrument to compare to, relying on cross-validation result will be more appropriate as physical meaningfulness would be harder to evaluate.

Figure 2 presents the bootstrap results of the four factors resolved from the NMF without normalization (left panels) and with normalization (right panels) as well as their diurnal patterns. The chemical composition of each factor is described by a fraction of species signal measured by the LCS. The numerical values of the bootstrap results can be found in the Supporting Information section 8. As seen in Figure 2a, factor 1 is dominated by CO (98.7%) and NO_2 (73.2%) with a small contribution of bin 0 (0.35−0.46 μm) (14.4%). Hence factor 1 is likely related to combustion.[78−80] Its diurnal pattern also exhibits peaks in the morning and the evening which reflects the emissions from traffic. Factor 2 is dominated by NO (83.1%) with a diurnal peak around noon time, similar to the diurnal pattern of NO measured from the LCS (Supporting Information section 2). Factor 3 is dominated by O_3 (99.1%) with a small contribution from NO (16.9%), NO_2 (18.6%), and bin 0 (0.35−0.46 μm) (6%). It peaks at 3 PM and again resembles the diurnal pattern of O_3 measured by the LCS. Lastly, factor 4 is composed of particles with bin 2 (0.66−1.0 μm) (115.6%) being the highest followed by bin 1 (0.46−0.66 μm) (112.4%) and bin 0 (0.35−0.46 μm) (76.7%). A previous study[49] that employed the NMF technique without normalization to study the sources of aerosol in Delhi, India, resolved three factors: a CO-dominated factor, an O_3-dominated factor, and a particle factor. Even though the settings of the current study and the previous study differ, the three factors identified by Hagan et al. and the four factors resolved from the non-normalized data set from the current study are all characterized by one or two species overly represented in each factor. On the other hand, the four factors resolved from the normalized data set from the current study are composed of more diverse species.

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

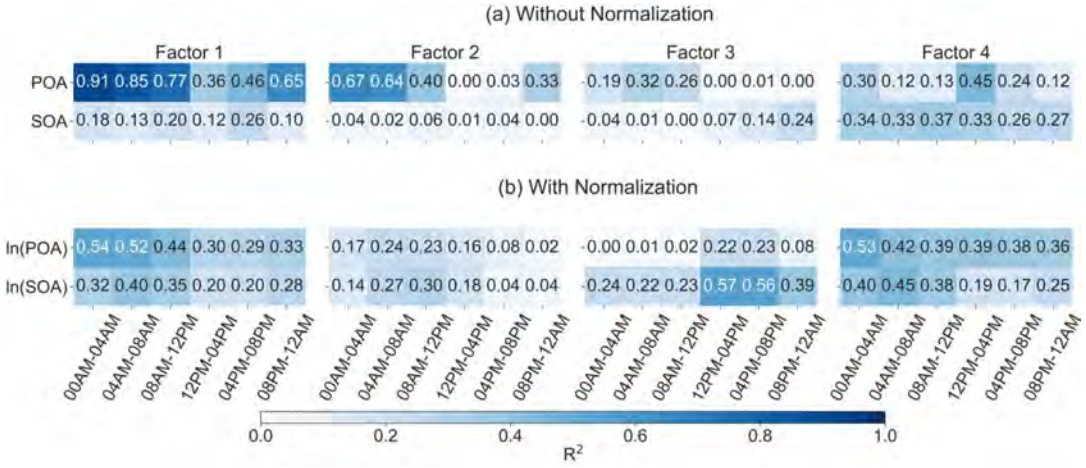


Figure 3. Correlation between each factor resolved from the NMF analysis and POA and SOA resolved from ACSM PMF analysis for two cases: (a) without normalization and (b) with normalization at different times of the day. The *x*-axis shows 24 h divided into 4 h segments.

Figure 2b illustrates the NMF result for the data set that has been normalized prior to running the NMF analysis. The first factor has a slight resemblance to factor 1 without normalization in that its dominant signals are CO (50.6%), NO_2 (51.9%), and bin 0 (50.9%). The difference is that with normalization, the solution contains other species such as NO (42.5%), bin 1 (21.1%), and bin 2 (14.2%) with less contribution from CO and NO_2 and more contribution from bin 0. The diurnal pattern is also different from the non-normalized data set's factor 1. Specifically, the non-normalized data set's factor 1 shows a clear peak during rush hours, whereas that of the normalized data set shows less distinctive peaks. This can be explained by the smaller fraction of CO being present in the normalized factor 1 (Supporting Information section 4). Factor 2 is a mixture of all gaseous species, CO (38.5%), NO (38.6%), NO_2 (36%), and O_3 (43.2%), with a small amount of bin 2 (9.5%) and its diurnal pattern resembles that of O_3. Factor 3 is like that of factor 3 from without normalization in that it is dominated by O_3 (56.8%) with a considerable contribution of bin 0 (32.8%), 1 (25.5%), and 2 (20.7%) and a minor contribution from NO (14.9%). However, the diurnal pattern is different from the non-normalized version with an earlier peak around 1 PM. This may be due to contribution of particle bins 0−2. Lastly, factor 4 has the strongest signal from particle bins 1 and 2 (both >50%) with a small contribution from bin 0 (16.4%), CO (10.6%), and NO_2 (9.6%) and a similar diurnal profile to the non-normalized version. Hence, except for factor 2, with and without normalization NMF results yield similar factors, including a combustion related factor, an O_3-dominant factor, and a particle-dominant factor.

It should be noted that normalization leads to more diversified profiles with multiple species mixed in for each factor. Moreover, not normalizing the data before running the NMF can lead to overestimation of species (e.g., bin 1 and 2 fractions are >100% in factor 3) whereas normalization prevents such overestimation. The non-normalized data may suffer from overestimation as each factor is represented by fewer LCS species. A similar overestimation is also reported in Hagan et al.[49] when non-normalized data are used for the NMF analysis.

Reproducibility in our analysis is determined by bootstrapping. If the standard deviation from 100 trials is high, it can be interpreted as a data set having a low reproducibility whereas low standard deviation indicates high reproducibility. The reproducibility of both without and with normalization is high with the average of all standard deviations being 0.4% for non-normalized and 0.9% for the normalized data set (Supporting Information section 8).

Linking Factors Resolved from NMF Analysis to NR-PM_1 Composition. To better understand the sources of the four factors resolved from NMF analysis, the correlation between these factors and the POA and SOA resolved from ACSM is plotted (Figure 3). The R^2 values from the ordinary least-squares fits between the factors are calculated every 4 h for the non-normalized (Figure 3a) and normalized data set (Figure 3b). In Figure 3, POA is the sum of HOA, COA, and BBOA, and SOA is the sum of LO-OOA and MO-OOA. More comprehensive plots, including the R^2 with parameters (total PM_1 mass and number concentrations) measured by the SMPS, chemical composition (Org, SO_4, NO_3, NH_4) measured by the ACSM, reactive gas species (e.g., O_3, NO_2, and O_x) from reference instruments, and HOA, COA, BBOA, MO-OOA, and LO-OOA from PMF analysis of the ACSM data can be found in the Supporting Information section 9.

As shown in Figure 3a, factor 1 (CO-dominant factor) has a substantial correlation with POA during the morning and evening rush hours and when the mixing depth is low at night. Such a relationship confirms that factor 1 is a combustion related factor, and the same diurnal pattern of aerosol related to combustion is observed in our prior studies at the same site and nearby sites.[35,37] Factor 2 (NO-dominant factor) resembles that of factor 1 and is again a factor related to primary emissions. Factor 3 (O_3-dominant factor) does not exhibit association with aerosol species and OA factors but illustrates a positive slope with O_3 (R^2 = 0.54 for 12 PM to 4 PM, Supporting Information sections 9 and 10). Lastly, factor 4 (particle-dominant factor) is distinctive from other factors in that it exhibits moderate to high correlations with inorganic aerosol such as SO_4, NO_3, and NH_4 (Supporting Information section 9). At the same time, it is moderately correlated with POA and Org (Supporting Information section 9). Thus, this factor can represent a mixture of inorganic and organic aerosol.

For the normalized data set sources (Figure 3b), factor 1 correlates well with POA at a similar time (i.e., morning and evening rush hours) as factor 1 without normalization.

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Therefore, it is also a combustion related factor. It has a higher correlation with inorganics such as NH_4, NO_3, SO_4, and SOA than the factor 1 without normalization (Supporting Information section 9). This may be due to factor 1 with normalization containing larger particle bins such as bins 1 and 2. For factor 2 (mixture of gases factor), it exhibits a negative relationship with PM_1, inorganics, and OA factors resolved from ACSM but a positive relationship with O_3 and O_x (Supporting Information sections 9 and 10). This resembles factor 3 resolved from the non-normalized data set. Factor 3 illustrates a substantial correlation with SOA from 12 PM to 8 PM. Hence, factor 3 can represent a daytime SOA factor. However, this SOA factor has almost no correlation with O_3 and O_x (Supporting Information section 10). SOA formed from photochemical production has been shown to have a high R^2 with O_3 and O_x[33,81,82] while SOA transported regionally shows a weak correlation with O_3.[83] Therefore, it is possible that factor 3 can be from long-range transport or entrainment of aged SOA from aloft.[84,85] Lastly, factor 4 is a nighttime SOA factor. Factor 4 has $R^2 > 0.3$ across all times with POA, which hints that the correlation may be driven by a similar diurnal pattern. Factor 4 demonstrates a moderate correlation with SOA from 00 AM to 08 AM. During that time, factor 4 shows an enhanced correlation (moderate to substantial) with PM_1, Org, and NO_3 (Supporting Information section 9). Previous studies conducted at the same site and in the southeastern U.S. show that nitrate radical oxidation of monoterpenes is an important source of SOA at night, which exhibits a strong correlation with particulate organic nitrate.[37,53,54]

In summary, the data set that employed normalization before running the NMF analysis can identify POA (factor 1), daytime SOA (factor 3), and nighttime SOA (factor 4), whereas the non-normalized data set is only able to identify factors related to POA. One advantage for the non-normalized data set is that its factor 1 correlation with POA is substantially higher, as it is exclusively composed of CO and NO_2. Also, if there is no research-grade instrument colocated with the LCS, the non-normalized data set may lead to a more intuitive interpretation of factors as each factor is composed of one or two species predominantly. However, a disadvantage of the non-normalized data set is that it may perform poorly on identifying different sources of aerosol as each factor's composition is not as diverse. As we show in this study, the normalized data set can resolve SOA and each factor is composed of more diverse species. Hence, testing the NMF technique with both normalized and non-normalized data sets is the recommendation of the current work.

There are several limitations of the study that would benefit from future research. First, the performance of the LCS, especially gas measurements (e.g., O_3 and NO), is not as accurate as the research-grade instruments and continued effort to improve their accuracy is needed. Second, LCS do not provide information on aerosol chemical composition, which leads to inherent difficulties in interpreting the factors.

Implications. This study illustrates that gaseous and particle species measured by LCS combined with the NMF technique can resolve different pollutant sources even in a moderately polluted urban area with complex emissions. The benefits and limitations of applying normalization to the LCS data set prior to employing the NMF analysis are evaluated by comparing results for both normalized and non-normalized data sets. Based on the analysis, the normalized data set resolves factors with contributions from more diverse input species, whereas the non-normalized data set shows only one or two input species having high contributions in each factor. Furthermore, a normalized data set can discern the SOA from POA for the particle factor, whereas the non-normalized data set can only identify multiple POA factors. However, the current work acknowledges that interpreting the normalized data set may be more complex if research-grade instruments are not available as the composition of each factor is more diverse. Taken together, this study demonstrates the new possibility of using LCS for aerosol source apportionment even in an area that is considerably affected by both anthropogenic and biogenic sources with complex chemistry, which typically requires the use of advanced instrumentation. More studies on the use of LCS for source apportionment in other environments are warranted. This work also shows a new possibility to do source apportionment at a wider scale even in areas that lack research-grade instruments and only have access to LCS. The continued growth in the application of LCS over finer spatial resolutions will open new opportunities to obtain new insights into various pollution sources and their atmospheric processing in the future.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.1c07005.

Detailed information about the potential role of RH on OPC measurements, the aerosol composition measured by the ACSM instrument, the normalization method, diurnal profiles of species measured by LCS, LCS performance evaluation, the different number of factors resolved from the NMF technique using the non-normalized data set, the mathematical formulation of PMF and NMF techniques, and comprehensive correlation plots for non-normalized and normalized data sets (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Nga Lee Ng − *School of Civil and Environmental Engineering, School of Chemical and Biomolecular Engineering, and School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0001-8460-4765; Email: ng@chbe.gatech.edu

Authors

Laura Hyesung Yang − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* Present Address: School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, United States; orcid.org/0000-0002-0057-7120

David H. Hagan − *QuantAQ, Inc., Somerville, Massachusetts 02143, United States;* orcid.org/0000-0001-5111-4671

Jean C. Rivera-Rios − *School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Makoto M. Kelp − *Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, United States;* orcid.org/0000-0003-0451-3439

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Eben S. Cross − *QuantAQ, Inc., Somerville, Massachusetts 02143, United States*
Yuyang Peng − *School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Jennifer Kaiser − *School of Civil and Environmental Engineering and School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*
Leah R. Williams − *Aerodyne Research, Inc., Billerica, Massachusetts 01821, United States*
Philip L. Croteau − *Aerodyne Research, Inc., Billerica, Massachusetts 01821, United States*
John T. Jayne − *Aerodyne Research, Inc., Billerica, Massachusetts 01821, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.1c07005

Notes

The authors declare the following competing financial interest(s): D.H.H. and E.S.C. are the co-founders of QuantAQ, Inc., which manufactures the low-cost air quality sensors used in this study. L.R.W., P.L.C., and J.T.J. are employed by Aerodyne Research, Inc., the manufacturer of the Aerosol Chemical Speciation Monitor used in this study. The remaining authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by NOAA Climate Program Office's Atmospheric Chemistry, Carbon Cycle, and Climate program, Grant Number NA21OAR4310221.

■ ABBREVIATIONS

ACSM	aerosol chemical speciation monitor
AMS	aerosol mass spectrometer
AE	auxiliary electrode
BBOA	biomass burning organic aerosol
CDCE	composition-dependent collection efficiency
CPC	condensation particle counter
COA	cooking organic aerosol
CV	cross-validation
HOA	hydrocarbon-like organic aerosol
LO-OOA	less-oxidized oxygenated organic aerosol
LCS	low-cost sensor
mV	millivolts
MO-OOA	more-oxidized oxygenated organic aerosol
NMF	non-negative matrix factorization
NR-PM_1	nonrefractory PM_1
N_{tot}	number concentration
OPC-N3	optical particle counter
OA	organic aerosol
OOA	oxygenated organic aerosol
$PM_{2.5}$	particulate matter smaller than 2.5 μm in diameter
PMF	positive matrix factorization
POA	primary organic aerosol
QuantAQ/Ari v200	QuantAQ/Aerodyne Research v200
RH	relative humidity
SMPS	scanning mobility particle sizer
SOA	secondary organic aerosol
WE	working electrode

■ REFERENCES

(1) Miller, D. J.; Actkinson, B.; Padilla, L.; Griffin, R. J.; Moore, K.; Lewis, P. G. T.; Gardner-Frolick, R.; Craft, E.; Portier, C. J.; Hamburg, S. P.; Alvarez, R. A. Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas. *Environ. Sci. Technol.* **2020**, *54* (4), 2133−2142.
(2) Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L. Deriving High-Resolution Urban Air Pollution Maps Using Mobile Sensor Nodes. *Pervasive Mob. Comput* **2015**, *16*, 268−285.
(3) Lewis, A.; Peltier, W. R.; von Schneidemesser, E. *Low-Cost Sensors for the Measurement of Atmospheric Composition: Overview of Topic and Future Applications*, World Meteorological Organization (WMO), 2018.
(4) Gupta, P.; Doraiswamy, P.; Levy, R.; Pikelnaya, O.; Maibach, J.; Feenstra, B.; Polidori, A.; Kiros, F.; Mills, K. C. Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations. *GeoHealth* **2018**, *2* (6), 172−181.
(5) Castell, N.; Dauge, F. R.; Schneider, P.; Vogt, M.; Lerner, U.; Fishbain, B.; Broday, D.; Bartonova, A. Can Commercial Low-Cost Sensor Platforms Contribute to Air Quality Monitoring and Exposure Estimates? *Environ. Int.* **2017**, *99*, 293−302.
(6) Munir, S.; Mayfield, M.; Coca, D.; Jubb, S. A.; Osammor, O. Analysing the Performance of Low-Cost Air Quality Sensors, Their Drivers, Relative Benefits and Calibration in Cities—a Case Study in Sheffield. *Environ. Monit. Assess* **2019**, *191* (2), 94.
(7) Tryner, J.; Phillips, M.; Quinn, C.; Neymark, G.; Wilson, A.; Jathar, S. H.; Carter, E.; Volckens, J. Design and Testing of a Low-Cost Sensor and Sampling Platform for Indoor Air Quality. *Build. Environ* **2021**, *206*, 108398.
(8) Chu, H.-J.; Ali, M. Z.; He, Y.-C. Spatial Calibration and PM2.5 Mapping of Low-Cost Air Quality Sensors. *Sci. Rep* **2020**, *10* (1), 22079.
(9) Cross, E. S.; Williams, L. R.; Lewis, D. K.; Magoon, G. R.; Onasch, T. B.; Kaminsky, M. L.; Worsnop, D. R.; Jayne, J. T. Use of Electrochemical Sensors for Measurement of Air Pollution: Correcting Interference Response and Validating Measurements. *Atmospheric Meas. Tech* **2017**, *10* (9), 3575−3588.
(10) Miskell, G.; Salmond, J. A.; Williams, D. E. Solution to the Problem of Calibration of Low-Cost Air Quality Measurement Sensors in Networks. *ACS Sens* **2018**, *3* (4), 832−843.
(11) van Zoest, V.; Osei, F. B.; Stein, A.; Hoek, G. Calibration of Low-Cost NO2 Sensors in an Urban Air Quality Network. *Atmos. Environ.* **2019**, *210*, 66−75.
(12) Cordero, J. M.; Borge, R.; Narros, A. Using Statistical Methods to Carry out in Field Calibrations of Low Cost Air Quality Sensors. *Sens. Actuators B Chem.* **2018**, *267*, 245−254.
(13) Hagan, D. H.; Isaacman-VanWertz, G.; Franklin, J. P.; Wallace, L. M. M.; Kocar, B. D.; Heald, C. L.; Kroll, J. H. Calibration and Assessment of Electrochemical Air Quality Sensors by Co-Location with Regulatory-Grade Instruments. *Atmospheric Meas. Tech* **2018**, *11* (1), 315−328.
(14) Zimmerman, N.; Presto, A. A.; Kumar, S. P. N.; Gu, J.; Hauryliuk, A.; Robinson, E. S.; Robinson, A. L.; Subramanian, R. A Machine Learning Calibration Model Using Random Forests to Improve Sensor Performance for Lower-Cost Air Quality Monitoring. *Atmospheric Meas. Tech.* **2018**, *11* (1), 291−313.
(15) Masson, N.; Piedrahita, R.; Hannigan, M. Quantification Method for Electrolytic Sensors in Long-Term Monitoring of Ambient Air Quality. *Sensors* **2015**, *15* (10), 27283−27302.
(16) Popoola, O. A. M.; Stewart, G. B.; Mead, M. I.; Jones, R. L. Development of a Baseline-Temperature Correction Methodology for Electrochemical Sensors and Its Implications for Long-Term Stability. *Atmos. Environ.* **2016**, *147*, 330−343.
(17) Smith, K. R.; Edwards, P. M.; Ivatt, P. D.; Lee, J. D.; Squires, F.; Dai, C.; Peltier, R. E.; Evans, M. J.; Sun, Y.; Lewis, A. C. An Improved

Low-Power Measurement of Ambient NO₂ and O₃ Combining Electrochemical Sensor Clusters and Machine Learning. *Atmospheric Meas. Tech* **2019**, *12* (2), 1325−1336.
(18) Tryner, J.; L'Orange, C.; Mehaffy, J.; Miller-Lionberg, D.; Hofstetter, J. C.; Wilson, A.; Volckens, J. Laboratory Evaluation of Low-Cost PurpleAir PM Monitors and in-Field Correction Using Co-Located Portable Filter Samplers. *Atmos. Environ.* **2020**, *220*, 117067.
(19) Tryner, J.; Mehaffy, J.; Miller-Lionberg, D.; Volckens, J. Effects of Aerosol Type and Simulated Aging on Performance of Low-Cost PM Sensors. *J. Aerosol Sci.* **2020**, *150*, 105654.
(20) Luo, L.; Zhang, Y.; Pearson, B.; Ling, Z.; Yu, H.; Fu, X. On the Security and Data Integrity of Low-Cost Sensor Networks for Air Quality Monitoring. *Sensors* **2018**, *18* (12), 4451.
(21) Williams, D. E. Low Cost Sensor Networks: How Do We Know the Data Are Reliable? *ACS Sens* **2019**, *4* (10), 2558−2565.
(22) Liu, X.; Jayaratne, R.; Thai, P.; Kuhn, T.; Zing, I.; Christensen, B.; Lamont, R.; Dunbabin, M.; Zhu, S.; Gao, J.; Wainwright, D.; Neale, D.; Kan, R.; Kirkwood, J.; Morawska, L. Low-Cost Sensors as an Alternative for Long-Term Air Quality Monitoring. *Environ. Res.* **2020**, *185*, 109438.
(23) Tai, A. P. K.; Mickley, L. J.; Jacob, D. J. Correlations between Fine Particulate Matter (PM2.5) and Meteorological Variables in the United States: Implications for the Sensitivity of PM2.5 to Climate Change. *Atmos. Environ.* **2010**, *44* (32), 3976−3984.
(24) Schleicher, N. J.; Norra, S.; Chai, F.; Chen, Y.; Wang, S.; Cen, K.; Yu, Y.; Stüben, D. Temporal Variability of Trace Metal Mobility of Urban Particulate Matter from Beijing − A Contribution to Health Impact Assessments of Aerosols. *Atmos. Environ.* **2011**, *45* (39), 7248−7265.
(25) Yuan, C.-S.; Lee, C.-G.; Liu, S.-H.; Chang, J.; Yuan, C.; Yang, H.-Y. Correlation of Atmospheric Visibility with Chemical Composition of Kaohsiung Aerosols. *Atmospheric Res.* **2006**, *82* (3), 663−679.
(26) Das, S. K.; Jayaraman, A. Long-Range Transportation of Anthropogenic Aerosols over Eastern Coastal Region of India: Investigation of Sources and Impact on Regional Climate Change. *Atmospheric Res.* **2012**, *118*, 68−83.
(27) Katul, G. G.; Grönholm, T.; Launiainen, S.; Vesala, T. The Effects of the Canopy Medium on Dry Deposition Velocities of Aerosol Particles in the Canopy Sub-Layer above Forested Ecosystems. *Atmos. Environ.* **2011**, *45* (5), 1203−1212.
(28) Paatero, P.; Tapper, U. Positive Matrix Factorization: A Non-Negative Factor Model with Optimal Utilization of Error Estimates of Data Values. *Environmetrics* **1994**, *5* (2), 111−126.
(29) Ulbrich, I. M.; Canagaratna, M. R.; Zhang, Q.; Worsnop, D. R.; Jimenez, J. L. Interpretation of Organic Components from Positive Matrix Factorization of Aerosol Mass Spectrometric Data. *Atmospheric Chem. Phys.* **2009**, *9* (9), 2891−2918.
(30) Hayes, P. L.; Ortega, A. M.; Cubison, M. J.; Froyd, K. D.; Zhao, Y.; Cliff, S. S.; Hu, W. W.; Toohey, D. W.; Flynn, J. H.; Lefer, B. L.; Grossberg, N.; Alvarez, S.; Rappengluck, B.; Taylor, J. W.; Allan, J. D.; Holloway, J. S.; Gilman, J. B.; Kuster, W. C.; de Gouw, J. A.; Massoli, P.; Zhang, X.; Liu, J.; Weber, R. J.; Corrigan, A. L.; Russell, L. M.; Isaacman, G.; Worton, D. R.; Kreisberg, N. M.; Goldstein, A. H.; Thalman, R.; Waxman, E. M.; Volkamer, R.; Lin, Y. H.; Surratt, J. D.; Kleindienst, T. E.; Offenberg, J. H.; Dusanter, S.; Griffith, S.; Stevens, P. S.; Brioude, J.; Angevine, W. M.; Jimenez, J. L. Organic Aerosol Composition and Sources in Pasadena, California, during the 2010 CalNex Campaign. *J. Geophys. Res. Atmospheres* **2013**, *118* (16), 9233−9257.
(31) Slowik, J. G.; Vlasenko, A.; McGuire, M.; Evans, G. J.; Abbatt, J. P. D. Simultaneous Factor Analysis of Organic Particle and Gas Mass Spectra: AMS and PTR-MS Measurements at an Urban Site. *Atmos Chem. Phys.* **2010**, *10* (4), 1969−1988.
(32) Dall'Osto, M.; Ovadnevaite, J.; Ceburnis, D.; Martin, D.; Healy, R. M.; O'Connor, I. P.; Kourtchev, I.; Sodeau, J. R.; Wenger, J. C.; O'Dowd, C. Characterization of Urban Aerosol in Cork City (Ireland) Using Aerosol Mass Spectrometry. *Atmospheric Chem. Phys.* **2013**, *13* (9), 4997−5015.
(33) Lanz, V. A.; Alfarra, M. R.; Baltensperger, U.; Buchmann, B.; Hueglin, C.; Prévôt, A. S. H. Source Apportionment of Submicron Organic Aerosols at an Urban Site by Factor Analytical Modelling of Aerosol Mass Spectra. *Atmospheric Chem. Phys.* **2007**, *7* (6), 1503−1522.
(34) 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. *Anal. Bioanal. Chem.* **2011**, *401* (10), 3045−3067.
(35) Xu, L.; Suresh, S.; Guo, H.; Weber, R. J.; Ng, N. L. Aerosol Characterization over the Southeastern United States Using High-Resolution Aerosol Mass Spectrometry: Spatial and Seasonal Variation of Aerosol Composition and Sources with a Focus on Organic Nitrates. *Atmospheric Chem. Phys.* **2015**, *15* (13), 7307−7336.
(36) 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 Chem. Phys.* **2010**, *10* (10), 4625−4641.
(37) Chen, Y.; Takeuchi, M.; Nah, T.; Xu, L.; Canagaratna, M. R.; Stark, H.; Baumann, K.; Canonaco, F.; Prévôt, A. S. H.; Huey, L. G.; Weber, R. J.; Ng, N. L. Chemical Characterization of Secondary Organic Aerosol at a Rural Site in the Southeastern US: Insights from Simultaneous High-Resolution Time-of-Flight Aerosol Mass Spectrometer (HR-ToF-AMS) and FIGAERO Chemical Ionization Mass Spectrometer (CIMS) Measurements. *Atmospheric Chem. Phys.* **2020**, *20* (14), 8421−8440.
(38) 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, E. 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.; Shimono, 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.
(39) Äijälä, M.; Heikkinen, L.; Fröhlich, R.; Canonaco, F.; Prévôt, A. S. H.; Junninen, H.; Petäjä, T.; Kulmala, M.; Worsnop, D.; Ehn, M. Resolving Anthropogenic Aerosol Pollution Types − Deconvolution and Exploratory Classification of Pollution Events. *Atmospheric Chem. Phys.* **2017**, *17* (4), 3165−3197.
(40) Raatikainen, T.; Vaattovaara, P.; Tiitta, P.; Miettinen, P.; Rautiainen, J.; Ehn, M.; Kulmala, M.; Laaksonen, A.; Worsnop, D. R. Physicochemical Properties and Origin of Organic Groups Detected in Boreal Forest Using an Aerosol Mass Spectrometer. *Atmospheric Chem. Phys.* **2010**, *10* (4), 2063−2077.
(41) Budisulistiorini, S. H.; Canagaratna, M. R.; Croteau, P. L.; Marth, W. J.; Baumann, K.; Edgerton, E. S.; Shaw, S. L.; Knipping, E. M.; Worsnop, D. R.; Jayne, J. T.; Gold, A.; Surratt, J. D. Real-Time Continuous Characterization of Secondary Organic Aerosol Derived from Isoprene Epoxydiols in Downtown Atlanta, Georgia, Using the Aerodyne Aerosol Chemical Speciation Monitor. *Environ. Sci. Technol.* **2013**, *47* (11), 5686−5694.
(42) Zhang, Y.; Tang, L.; Yu, H.; Wang, Z.; Sun, Y.; Qin, W.; Chen, W.; Chen, C.; Ding, A.; Wu, J.; Ge, S.; Chen, C.; Zhou, H. Chemical Composition, Sources and Evolution Processes of Aerosol at an Urban Site in Yangtze River Delta, China during Wintertime. *Atmos. Environ.* **2015**, *123*, 339−349.

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(43) Sun, Y.; Wang, Z.; Dong, H.; Yang, T.; Li, J.; Pan, X.; Chen, P.; Jayne, J. T. Characterization of Summer Organic and Inorganic Aerosols in Beijing, China with an Aerosol Chemical Speciation Monitor. *Atmos. Environ.* **2012**, *51*, 250−259.
(44) Crenn, V.; Sciare, J.; Croteau, P. L.; Verlhac, S.; Fröhlich, R.; Belis, C. A.; Aas, W.; Äijälä, M.; Alastuey, A.; Artiñano, B.; Baisnée, D.; Bonnaire, N.; Bressi, M.; Canagaratna, M.; Canonaco, F.; Carbone, C.; Cavalli, F.; Coz, E.; Cubison, M. J.; Esser-Gietl, J. K.; Green, D. C.; Gros, V.; Heikkinen, L.; Herrmann, H.; Lunder, C.; Minguillón, M. C.; Močnik, G.; O'Dowd, C. D.; Ovadnevaite, J.; Petit, J.-E.; Petralia, E.; Poulain, L.; Priestman, M.; Riffault, V.; Ripoll, A.; Sarda-Estève, R.; Slowik, J. G.; Setyan, A.; Wiedensohler, A.; Baltensperger, U.; Prévôt, A. S. H.; Jayne, J. T.; Favez, O. ACTRIS ACSM Intercomparison − Part 1: Reproducibility of Concentration and Fragment Results from 13 Individual Quadrupole Aerosol Chemical Speciation Monitors (Q-ACSM) and Consistency with Co-Located Instruments. *Atmospheric Meas. Tech* **2015**, *8* (12), 5063−5087.
(45) Fröhlich, R.; Crenn, V.; Setyan, A.; Belis, C. A.; Canonaco, F.; Favez, O.; Riffault, V.; Slowik, J. G.; Aas, W.; Aijälä, M.; Alastuey, A.; Artiñano, B.; Bonnaire, N.; Bozzetti, C.; Bressi, M.; Carbone, C.; Coz, E.; Croteau, P. L.; Cubison, M. J.; Esser-Gietl, J. K.; Green, D. C.; Gros, V.; Heikkinen, L.; Herrmann, H.; Jayne, J. T.; Lunder, C. R.; Minguillón, M. C.; Močnik, G.; O'Dowd, C. D.; Ovadnevaite, J.; Petralia, E.; Poulain, L.; Priestman, M.; Ripoll, A.; Sarda-Estève, R.; Wiedensohler, A.; Baltensperger, U.; Sciare, J.; Prévôt, A. S. H. ACTRIS ACSM Intercomparison − Part 2: Intercomparison of ME-2 Organic Source Apportionment Results from 15 Individual, Co-Located Aerosol Mass Spectrometers. *Atmospheric Meas. Tech* **2015**, *8* (6), 2555−2576.
(46) 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.
(47) Heikkinen, L.; Äijälä, M.; Daellenbach, K. R.; Chen, G.; Garmash, O.; Aliaga, D.; Graeffe, F.; Räty, 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 Chem. Phys.* **2021**, *21* (13), 10081−10109.
(48) Pope, F. D.; Gatari, M.; Ng'ang'a, D.; Poynter, A.; Blake, R. Airborne Particulate Matter Monitoring in Kenya Using Calibrated Low-Cost Sensors. *Atmospheric Chem. Phys.* **2018**, *18* (20), 15403−15418.
(49) Hagan, D. H.; Gani, S.; Bhandari, S.; Patel, K.; Habib, G.; Apte, J. S.; Hildebrandt Ruiz, L.; Kroll, J. H. Inferring Aerosol Sources from Low-Cost Air Quality Sensor Measurements: A Case Study in Delhi, India. *Environ. Sci. Technol. Lett.* **2019**, *6* (8), 467−472.
(50) Bousiotis, D.; Singh, A.; Haugen, M.; Beddows, D. C. S.; Diez, S.; Murphy, K. L.; Edwards, P. M.; Boies, A.; Harrison, R. M.; Pope, F. D. Assessing the Sources of Particles at an Urban Background Site Using Both Regulatory Instruments and Low-Cost Sensors − a Comparative Study. *Atmospheric Meas. Tech* **2021**, *14* (6), 4139−4155.
(51) Evans, C.; Hardin, J.; Stoebel, D. M. Selecting Between-Sample RNA-Seq Normalization Methods from the Perspective of Their Assumptions. *Brief. Bioinform* **2018**, *19* (5), 776−792.
(52) Hsu, L. L.; Culhane, A. C. Impact of Data Preprocessing on Integrative Matrix Factorization of Single Cell Data. *Front. Oncol* **2020**, *10*, 973.
(53) Xu, L.; Guo, H.; Boyd, C. M.; Klein, M.; Bougiatioti, A.; Cerully, K. M.; Hite, J. R.; Isaacman-VanWertz, G.; Kreisberg, N. M.; Knote, C.; Olson, K.; Koss, A.; Goldstein, A. H.; Hering, S. V.; de Gouw, J.; Baumann, K.; Lee, S.-H.; Nenes, A.; Weber, R. J.; Ng, N. L. Effects of Anthropogenic Emissions on Aerosol Formation from Isoprene and Monoterpenes in the Southeastern United States. *Proc. Natl. Acad. Sci.* **2015**, *112* (1), 37−42.
(54) Joo, T.; Chen, Y.; Xu, W.; Croteau, P.; Canagaratna, M. R.; Gao, D.; Guo, H.; Saavedra, G.; Kim, S. S.; Sun, Y.; Weber, R.; Jayne, J.; Ng, N. L. Evaluation of a New Aerosol Chemical Speciation Monitor (ACSM) System at an Urban Site in Atlanta, GA: The Use of Capture Vaporizer and PM2.5 Inlet. *ACS Earth Space Chem.* **2021**, *5* (10), 2565−2576.
(55) Matthew, B. M.; Middlebrook, A. M.; Onasch, T. B. Collection Efficiencies in an Aerodyne Aerosol Mass Spectrometer as a Function of Particle Phase for Laboratory Generated Aerosols. *Aerosol Sci. Technol.* **2008**, *42* (11), 884−898.
(56) Baron, R.; Saffell, J. Amperometric Gas Sensors as a Low Cost Emerging Technology Platform for Air Quality Monitoring Applications: A Review. *ACS Sens* **2017**, *2* (11), 1553−1566.
(57) Mead, M. I.; Popoola, O. A. M.; Stewart, G. B.; Landshoff, P.; Calleja, M.; Hayes, M.; Baldovi, J. J.; McLeod, M. W.; Hodgson, T. F.; Dicks, J.; Lewis, A.; Cohen, J.; Baron, R.; Saffell, J. R.; Jones, R. L. The Use of Electrochemical Sensors for Monitoring Urban Air Quality in Low-Cost, High-Density Networks. *Atmos. Environ.* **2013**, *70*, 186−203.
(58) Mueller, M.; Meyer, J.; Hueglin, C. Design of an Ozone and Nitrogen Dioxide Sensor Unit and Its Long-Term Operation within a Sensor Network in the City of Zurich. *Atmospheric Meas. Tech* **2017**, *10* (10), 3783−3799.
(59) Spinelle, L.; Gerboles, M.; Villani, M. G.; Aleixandre, M.; Bonavitacola, F. Field Calibration of a Cluster of Low-Cost Available Sensors for Air Quality Monitoring. Part A: Ozone and Nitrogen Dioxide. *Sens. Actuators B Chem.* **2015**, *215*, 249−257.
(60) Waskom, M. Seaborn: Statistical Data Visualization. *J. Open Source Softw* **2021**, *6* (60), 3021.
(61) McKinney, W. Data Structures for Statistical Computing in Python. *Python in Science Conference*, Austin, TX, 2010; pp 56−61, DOI: 10.25080/Majora-92bf1922-00a.
(62) McKinney, W. *pandas: A Foundational Python Library for Data Analysis and Statistics*, 2011.
(63) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13* (2), 22−30.
(64) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(65) Lee, D. D.; Seung, H. S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **1999**, *401* (6755), 788−791.
(66) Pascual-Montano, A.; Carmona-Saez, P.; Chagoyen, M.; Tirado, F.; Carazo, J. M.; Pascual-Marqui, R. D. BioNMF: A Versatile Tool for Non-Negative Matrix Factorization in Biology. *BMC Bioinformatics* **2006**, *7* (1), 366.
(67) Yan, M.; Yang, X.; Hang, W.; Xia, Y. Determining the Number of Factors for Non-Negative Matrix and Its Application in Source Apportionment of Air Pollution in Singapore. *Stoch. Environ. Res. Risk Assess* **2019**, *33* (4−6), 1175−1186.
(68) Kfoury, A.; Ledoux, F.; Roche, C.; Delmaire, G.; Roussel, G.; Courcot, D. PM2.5 Source Apportionment in a French Urban Coastal Site under Steelworks Emission Influences Using Constrained Non-Negative Matrix Factorization Receptor Model. *J. Environ. Sci.* **2016**, *40*, 114−128.
(69) Reff, A.; Eberly, S. I.; Bhave, P. V. Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods. *J. Air Waste Manag. Assoc* **2007**, *57* (2), 146−154.
(70) Paatero, P. Least Squares Formulation of Robust Non-Negative Factor Analysis. *Chemom. Intell. Lab. Syst* **1997**, *37* (1), 23−35.
(71) Li, H.; Guensler, R.; Ogle, J.; Jun, J.; Elango, V. V. Descriptive Analysis of Long-Distance Travel by Personal Vehicles Using 2004 Commute Atlanta Data. In *Transportation Research Board 86th Annual Meeting*, 2007; p 07-1688.

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(72) Hagan, D. H.; Kroll, J. H. Assessing the Accuracy of Low-Cost Optical Particle Sensors Using a Physics-Based Approach. *Atmospheric Meas. Tech* **2020**, *13* (11), 6343−6355.
(73) Xu, L.; Guo, H.; Weber, R. J.; Ng, N. L. Chemical Characterization of Water-Soluble Organic Aerosol in Contrasting Rural and Urban Environments in the Southeastern United States. *Environ. Sci. Technol.* **2017**, *51* (1), 78−88.
(74) Xu, L.; Pye, H. O. T.; He, J.; Chen, Y.; Murphy, B. N.; Ng, N. L. Experimental and Model Estimates of the Contributions from Biogenic Monoterpenes and Sesquiterpenes to Secondary Organic Aerosol in the Southeastern United States. *Atmospheric Chem. Phys.* **2018**, *18* (17), 12613−12637.
(75) Efron, B.; Tibshirani, R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statist. Sci.* **1986**, *1* (1), 54−75.
(76) Bro, R.; Kjeldahl, K.; Smilde, A. K.; Kiers, H. A. L. Cross-Validation of Component Models: A Critical Look at Current Methods. *Anal. Bioanal. Chem.* **2008**, *390* (5), 1241−1251.
(77) Paatero, P.; Eberly, S.; Brown, S. G.; Norris, G. A. Methods for Estimating Uncertainty in Factor Analytic Solutions. *Atmospheric Meas. Tech* **2014**, *7* (3), 781−797.
(78) Kelp, M.; Gould, T.; Austin, E.; Marshall, J. D.; Yost, M.; Simpson, C.; Larson, T. Sensitivity Analysis of Area-Wide, Mobile Source Emission Factors to High-Emitter Vehicles in Los Angeles. *Atmos. Environ.* **2020**, *223*, 117212.
(79) Pachon, J. E.; Balachandran, S.; Hu, Y.; Mulholland, J. A.; Darrow, L. A.; Sarnat, J. A.; Tolbert, P. E.; Russell, A. G. Development of Outcome-Based, Multipollutant Mobile Source Indicators. *J. Air Waste Manag. Assoc* **2012**, *62* (4), 431−442.
(80) Park, S. S.; Vijayan, A.; Mara, S. L.; Herner, J. D. Investigating the Real-World Emission Characteristics of Light-Duty Gasoline Vehicles and Their Relationship to Local Socioeconomic Conditions in Three Communities in Los Angeles, California. *J. Air Waste Manag. Assoc* **2016**, *66* (10), 1031−1044.
(81) Volkamer, R.; Jimenez, J. L.; San Martini, F.; Dzepina, K.; Zhang, Q.; Salcedo, D.; Molina, L. T.; Worsnop, D. R.; Molina, M. J. Secondary Organic Aerosol Formation from Anthropogenic Air Pollution: Rapid and Higher than Expected. *Geophys. Res. Lett.* **2006**, *33*, L17811.
(82) Herndon, S. C.; Onasch, T. B.; Wood, E. C.; Kroll, J. H.; Canagaratna, M. R.; Jayne, J. T.; Zavala, M. A.; Knighton, W. B.; Mazzoleni, C.; Dubey, M. K.; Ulbrich, I. M.; Jimenez, J. L.; Seila, R.; de Gouw, J. A.; de Foy, B.; Fast, J.; Molina, L. T.; Kolb, C. E.; Worsnop, D. R. Correlation of Secondary Organic Aerosol with Odd Oxygen in Mexico City. *Geophys. Res. Lett.* **2008**, *35*, L15804.
(83) Zhang, Q.; Worsnop, D. R.; Canagaratna, M. R.; Jimenez, J. L. Hydrocarbon-like and Oxygenated Organic Aerosols in Pittsburgh: Insights into Sources and Processes of Organic Aerosols. *Atmos Chem. Phys.* **2005**, *5* (12), 3289−3311.
(84) Nagori, J.; Janssen, R. H. H.; Fry, J. L.; Krol, M.; Jimenez, J. L.; Hu, W.; Vila-Guerau de Arellano, J. Biogenic Emissions and Land−Atmosphere Interactions as Drivers of the Daytime Evolution of Secondary Organic Aerosol in the Southeastern US. *Atmos Chem. Phys.* **2019**, *19* (2), 701−729.
(85) Chen, Y.; Guo, H.; Nah, T.; Tanner, D. J.; Sullivan, A. P.; Takeuchi, M.; Gao, Z.; Vasilakos, P.; Russell, A. G.; Baumann, K.; Huey, L. G.; Weber, R. J.; Ng, N. L. Low-Molecular-Weight Carboxylic Acids in the Southeastern U.S.: Formation, Partitioning, and Implications for Organic Aerosol Aging. *Environ. Sci. Technol.* **2021**, *55* (10), 6688−6699.

https://doi.org/10.1021/acs.est.1c07005
Environ. Sci. Technol. XXXX, XXX, XXX−XXX