**ABSTRACT**

Positive Matrix Factorization (PMF) is a commonly used receptor model for source apportionment of PM$_{2.5}$. However, PMF results often retrieve an individual factor mainly composed of secondary aerosols, making it difficult to link with primary emission sources and formulate effective air pollution control strategies. To overcome this limitation, we employed a two-stage PMF modeling approach with adjustments of the species weighting, which was fused with a robust regression model to better characterize the sources of PM$_{2.5}$ secondary aerosols. Additionally, organic molecular tracers were incorporated into PMF for source identification. A field campaign was conducted between May and December 2021 in Taichung, Taiwan. An improved PMF model was utilized to resolve the multiple time resolution data of 3-h online and 24-h offline measurements of PM$_{2.5}$ compositions. Retrieved factors from PMF were averaged over 24-h intervals and then applied in robust regression analysis to re- apportion the contributions. Comparing with conventional PMF, downweighting the secondary aerosol-related species in the model was more effective in linking them to primary emission sources. The results from the fusion model showed that the majority of secondary aerosols (sum of secondary aerosol-related species = 2.67 µg m$^{-3}$) within three hours were mainly contributed by oil combustion, while the largest contributor of secondary aerosols (1.65 µg m$^{-3}$) over 24 hours was industry, highlighting the need for regulation of these two sources based on various temporal scales. The developed fusion strategy of two-stage PMF and robust regression provided refined results and can aid in the management of PM$_{2.5}$.

**Keywords:** Secondary air pollution, Online measurements, Receptor model, Constraint

**1 INTRODUCTION**

Fine particulate matter (PM$_{2.5}$, particulate matter with an aerodynamic diameter of less than 2.5 µm) has been demonstrated to cause multiple adverse health outcomes in many epidemiological studies (Ghazi et al., 2022; Savouré et al., 2021; Tamayo-Ortiz et al., 2021; Wang et al., 2021; Zhao et al., 2022). To facilitate control management, identifying and quantifying the potential origins of PM$_{2.5}$ is essential. The receptor model of Positive Matrix Factorization (PMF) is widely used for the source apportionment of PM$_{2.5}$, providing scientific information on air pollution sources...
PMF decomposes the temporally measured particle composition matrix into factor profiles and contributions based on correlations among speciated data. Air pollution types can be recognized through loadings of marker species in retrieved profiles. The contributions of air pollutants to each factor are also quantitatively estimated (Norriss et al., 2014; Paatero and Tapper, 1994).

The speciated data of elements, water-soluble ions and carbonaceous components are extensively used for PMF modeling of PM$_{2.5}$. Nevertheless, an individual factor dominated by secondary aerosols was generally retrieved from PMF results and accounted for a significant part of PM$_{2.5}$ mass in many previous studies (Chi et al., 2019; Laiman et al., 2022; Li et al., 2020; Liao et al., 2017; Oduber et al., 2021; Park et al., 2019; Sharma et al., 2016; Taghvaei et al., 2018). Since the secondary aerosol-related species (e.g., sulfate, nitrate) were mainly distributed in one factor, it is difficult to recognize their associated primary emission sources and formulate effective PM$_{2.5}$ control strategies for secondary aerosols. This highlights the limitation of relying solely on PMF and the need to integrate with additional modeling approaches to apportion secondary aerosols.

To facilitate the PMF model analysis, a speciated data matrix and an associated uncertainty matrix should be provided. The uncertainty is determined based on the measurements and signal-to-noise ratio (S/N), which is utilized to assign the species categorization in PMF (namely strong, weak and bad). If a low S/N ratio is calculated, the species is classified as ‘weak’ and a high level of uncertainty should be given (Heo et al., 2009; Norriss et al., 2014), resulting in the weight of the species being reduced in the computation and avoiding a large influence on the model solution. In addition, the user-specified species of ‘total variable’, which represents the target being apportioned, should also be set as ‘weak’, thus the retrieved factors are mainly characterized and constituted by the input ‘strong’ species. In previous studies, various species (e.g., organic carbon or dust-related species) have been assigned as the ‘total variable’ for apportionment depending on the research purposes (Altuwayjiri et al., 2021; Hettiyadura et al., 2018; Lu et al., 2018). Thus, in the case of identifying the sources of secondary aerosol-related species, which are the targets to be apportioned, a ‘weak’ setting should be considered accordingly.

Comparing with secondary inorganic aerosol (SIA, which mainly includes the species of sulfate, nitrate and ammonium), the origins of secondary organic aerosol (SOA) could be more complex because of the gas-to-particle conversion process of a wide variety of volatile organic compounds (VOCs) through oxidation (Hallquist et al., 2009; Kleindienst et al., 2007), which can be affected by diverse factors, such as the concentrations of precursors and reacted chemicals (e.g., hydroxyl radicals) as reported in the literatures (Liu et al., 2023; Srivastava et al., 2022). Furthermore, SOA can account for more than 30% of PM$_{2.5}$ mass (Mancilla et al., 2015; Wang et al., 2017), indicating the need to identify the corresponding primary sources. Organic molecular tracers, which are photochemically reacted and derived from various VOC precursors, were utilized as SOA tracers in PMF modeling to characterize the primary emission sources of secondary organic carbon (SOC) (Feng et al., 2013; Hopke, 2016; Hu et al., 2010; Lanzafame et al., 2021; Zhang et al., 2009). In contrast to the PMF results without using tracers, the SOA sources can be additionally identified and separated by adding tracers. For example, as evidenced by Wang et al. (2012), the biogenic emission source cannot be highlighted by PMF when applying only elements, water-soluble ions, and carbonaceous components; however, it was additionally extracted as an isoprene SOA factor by including 2-methylerythritol (2-MT, a tracer of isoprene contributed from biogenic sources). Besides, more number of factor solutions were obtained, demonstrating that a refined PMF solution can be achieved by incorporating organic tracers (Wang et al., 2017; Wang et al., 2012). Previous studies also showed that the incorporation of SOA tracers in PMF can help in estimating the SOC contributions from different sources (Feng et al., 2013; Hu et al., 2010).

This study aimed to investigate the contribution of primary emission factors to secondary aerosols by utilizing both online and offline speciated PM$_{2.5}$ data in a developed fusion strategy. To achieve this objective, the following steps were taken: 1) characterize the contribution of secondary organic sources by analyzing organic molecular tracers and combine them with online speciated PM$_{2.5}$ data in PMF modeling; 2) propose a two-stage PMF modeling approach to apportion the origins of secondary aerosol-related species from retrieved factors; 3) investigate and quantify the associations between PMF-derived factors to facilitate the re-apportionment of secondary aerosols using a robust regression model.
2 METHODS

2.1 Speciated PM$_{2.5}$ Data Collections

The PM$_{2.5}$ components used in this study were collected at Zhongming air quality monitoring station (AQS) (24°9′7″N, 120°38′28″E), which is set up on the rooftop of 4th floor of Taichung Municipal Taichung Special Education School (Fig. 1) in an urban area of Taichung, Taiwan. Zhongming AQS’s sample inlet is 17.5 meters above the ground and has a distance of 49 meters from a major road. The AQS is equipped with an Ambient/Fence-Line Multi-Metals Monitor (Xact 625i, Cooper Environmental Services, OR, USA) and an In-situ Gas and Aerosol Composition monitor (IGAC S-611, Zhang Jia, Taiwan), which can provide hourly PM$_{2.5}$ elemental and ionic compositions, respectively. For detailed information on these semi-continuous instruments, please refer to previous studies (Huang et al., 2022; Young et al., 2022). Hourly PM$_{2.5}$ mass concentration is monitored using a BAM-1020 (Met One Instruments, Inc., OR, USA) provided by Taiwan Environmental Protection Administration.

In addition to hourly speciated PM$_{2.5}$ measurements, the 24-h integrated filter samples were simultaneously collected using Automatic Filter-Changing Samplers (AFCS, PNS 16T-3.1, Comde-Derenda GmbH, Germany) (Siudek and Ruczyńska, 2021; Yatkin et al., 2020). The PM$_{2.5}$ samples were collected on 47 mm quartz-fiber filters at a constant flow of 2.3 m$^3$ h$^{-1}$. The filters were pre-baked at 900°C for 4.5 hours before ambient sampling. The AFCS automatically changed the filters every 24 hours until all sampling tasks ended. The sampled filters were then transported from AQS to the laboratory and stored at −15°C in a refrigerator till the implementation of chemical analysis. To evaluate potential contamination during the sampling periods, field blanks were collected and analyzed for quality control and blank subtraction. The field campaign was conducted in four periods (May, August to September, October, and December) in 2021.

![Fig. 1. Location of Zhongming AQS and surrounding land use characteristics.](image-url)
2.2 Chemical Analysis for Organic Molecular Tracers and Carbonaceous Components

To identify the organic aerosol sources of PMF factors, four organic molecular tracers of primary organic aerosol (POA) and SOA were selected and chemically analyzed utilizing the collected filter samples. For POA, levoglucosan was chosen as the marker species of biomass burning (Bhattarai et al., 2019; Feng et al., 2013), which is commonly identified in the study area because of seasonal agricultural waste burning (Cheng et al., 2009; Laiman et al., 2022). For SOA, phthalic acid (PA) and isophthalic acid (IPA) were analyzed as tracers of naphthalene, which can be contributed from vehicular and industrial emissions; 2,3-dihydroxy-4-oxopentanoic acid (DHOPA), the tracer of aromatic compounds (e.g., toluene and xylene), was selected for the identification of anthropogenic emission (e.g., solvent use and industrial emission); 2-methylerythritol, the tracer of isoprene, was selected as the marker of biogenic emission (Al-Naiema and Stone, 2017; Feng et al., 2013; Hu et al., 2010; Kleindienst et al., 2007).

A Waters Ultra Performance Liquid Chromatography I-Class PLUS System (Waters, Milford, USA) coupled with a Waters Xevo TQ-XS tandem mass spectrometer (UPLC-MS/MS) was used to determine the organic tracers (Albinet et al., 2019). In sample pretreatment, 1000 ng of each internal standard (Table S1) in 0.1 mL methanol (MeOH) was spiked on the filter, and 10 mL MeOH was added as the solvent to dissolve the analytes. Then a vortex machine was used to shake the extraction bottle for 1.5 minutes and ultrasonicate for 20 minutes. Centrifugation was performed to remove the filter pieces. The extract was then filtered using a PTFE filter disc with 0.22 µm pore size. The filtered extract was preserved in the amber vial for the following chromatography analysis with UPLC-MS/MS. Method detection limits were calculated from the tripled standard deviations of the laboratory blank values. The carbonaceous components of organic carbon (OC) and elemental carbon (EC) were analyzed using the collected filter samples and a semi-continuous OC/EC Analyzer (Model-4, Sunset Laboratory, OR, USA) based on the thermal optical transmittance protocol with the NIOSH 5040 method.

2.3 Estimation of Primary and Secondary Organic Carbon

To estimate SOC, a minimum R squared method was utilized (Wu and Yu, 2016). This method uses the time series measurements of OC and EC to approximate the suitable ratio of primary OC/EC ((OC/EC)$_{pri}$). Sequential values of (OC/EC)$_{pri}$ are tested and set in the following equation to calculate SOC:

\[
SOC = OC_{total} - POC = OC_{total} - (OC/EC)_{pri} \times EC
\]  

Next, the $R^2$ values for EC and SOC are computed. The best (OC/EC)$_{pri}$ ratio is obtained where the minimum $R^2$ of EC and SOC presents. Since EC is contributed from primary sources exclusively, the lowest correlated SOC yields the most reasonable ratio estimate of (OC/EC)$_{pri}$, which is used to estimate SOC and POC in the equation above. The estimated SOC was chosen as the target for apportionment because it accounts for a significant fraction of SOA (Chou et al., 2017; Turpin and Lim, 2001).

2.4 The Procedures of Fusion Strategy for Source Apportionment

In this study, a fusion strategy was developed to apportion four PM$_{2.5}$ secondary aerosol-related species, which were the measured SIA species of sulfate (SO$_4^{2-}$), nitrate (NO$_3^-$) and ammonium (NH$_4^+$), and the estimated SOC (calculated via OC and EC described in Section 2.3). First, the collected speciated PM$_{2.5}$ data were applied to PMF modeling consecutively, but with distinct parameter settings. Next, a robust regression model was utilized to investigate and quantify the associations between secondary and primary feature PMF factors. For an illustration of the modeling procedure, please refer to Fig. 2.

2.4.1 PMF modeling with multiple time resolution data and constraints

To deconvolute the multiple time resolution speciated data (online monitoring measurements...
Fig. 2. Procedure of fusion strategy using two-stage PMF and robust regression model.

and 24-h filter samples) collected in this study, an improved PMF model was employed (Liao et al., 2013, 2017; Ogulei et al., 2005; Zhou et al., 2004):

\[ x_{ij} = \frac{1}{t_{s2} - t_{s1}} \sum_{i=t_{s1}}^{t_{s2}} \left( \sum_{k=1}^{n_k} g_{ik} f_{kj} \right) + e_{ij} \]  

(2)

where \( x_{ij} \) denotes the mass concentration matrix of the \( j \)th species in the \( s \)th sample; \( f_{kj} \) is the profile matrix of \( j \)th species in the \( k \)th factor; \( g_{ik} \) denotes the contribution matrix of the \( k \)th factor during the time units for the \( i \)th sample, and \( e_{ij} \) represents the residual for the \( j \)th species in the \( s \)th sample. Variables \( t_{s1} \) and \( t_{s2} \) are the start and end time, respectively, representing the number of time units and the highest time resolution interval in the input data. When \( t_{s1} \) is equal to \( t_{s2} \), the equation yields the same as the typical PMF mass balance equation (Norris et al., 2014; Paatero and Tapper, 1994):

\[ x_{ij} = \sum_{k=1}^{n_k} g_{ik} f_{kj} + e_{ij} \]  

(3)

With the application of the improved PMF model, the low time resolution data can be adequately utilized in the computation to identify the factor characterization without smoothing out the temporal variation of high time resolution data. The optimal model solution was selected by minimizing the objective function \( Q \):

\[ Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left( \frac{e_{ij}}{u_{ij}} \right)^2 \]  

(4)

where \( u_{ij} \) denotes the uncertainty associated with the \( j \)th species in the \( i \)th sample. As revealed by the equation, when the uncertainty \( u_{ij} \) increases, the objective function \( Q \) decreases, therefore having less influence on the model solution. The preprocessing of measured concentrations and calculation of uncertainty and S/N ratios of PM2.5 components followed the U.S. Environmental Protection Agency PMF 5.0 user guide (Norris et al., 2014). Two indicators, the maximum individual
column mean (IM) and the maximum individual column standard deviation (IS), were applied to search for the proper number of factors (Lee et al., 1999; Liao et al., 2013).

To apportion the SIA species and SOC contributed by primary emission factors, a two-stage PMF modeling was developed. In the first stage, all PM$_{2.5}$ components were put into the PMF model. In addition, the SIA species and SOC were set as ‘weak’ and the estimated uncertainty values were tripled to reduce their corresponding influence in modeling. The categorizations of the remaining species were set as ‘strong’ (i.e., without adjusting for uncertainty) because of their high S/N ratios. The first stage aims to give a solution dominantly featured by primary species, enabling improved association with secondary aerosol-related species. Furthermore, since the formation of secondary aerosols could require a longer time (Chuang et al., 2022), secondary aerosol-related species may not be fully explained by primary feature factors retrieved based on modeling interval of PMF inputs. Hence, a second-stage PMF coupling with another statistical approach was proposed to re-apportion secondary aerosols over an extended time interval. In the second-stage PMF, the unexplained secondary aerosol-related species were assigned to a differentiated factor through the following model parameter settings. Firstly, all factors were constrained based on the whole profile features retrieved from the first stage to preserve the characteristics of the resolved factors. In addition, the number of factor solution was increased by one compared to the first stage, with the re-categorization of SIA species and SOC from ‘weak’ to ‘strong’. Ultimately, the obtained factors, which facilitated the re-apportionment of secondary aerosols, were further assessed using a regression analysis (Section 2.4.2).

A flexible script-based program, Multilinear Engine version 2 (ME-2), was employed to execute the improved PMF and two-stage modeling approach (Liao et al., 2013; Paatero, 1999; Ramadan et al., 2003). The constraint procedure was imposed with “fkey” commands in the script of ME-2 to force the values of each species to fall within the predefined limits, which were set as 1% of fluctuation ranges of profile values retrieved from the first-stage PMF. In addition, to achieve a more interpretable result of apportioned PM$_{2.5}$ contributions, a ‘missing mass’ variable was input instead of including PM$_{2.5}$ mass as the target for apportionment in modeling (Huang et al., 2022; Larson et al., 2006; Wu et al., 2007). The ‘missing mass’ represents the concentration differences between the PM$_{2.5}$ and the sum of all measured species, except for organic molecular tracers, to avoid double counting. This additional variable characterized the PM$_{2.5}$ component part that was not measured in this study and constrained the sum of all species in each factor to be equal to the apportioned PM$_{2.5}$. The 100% of the sum of species mass fractions of each factor was computed by dividing the total species concentration.

2.4.2 Robust regression modeling with PMF-retrieved factors

To investigate and quantify the associations between the secondary and primary feature factors retrieved from PMF, a statistical approach of robust regression, which iteratively reweights the input values to reduce contamination from outliers (i.e., influential observations) in the model, was applied (Chen et al., 2012; Liao et al., 2017):

$$\text{[Secondary feature factor contribution]} = b_0 + b_1 \times \text{[Primary feature factor 1 contribution]} + \ldots + b_n \times \text{[Primary feature factor n contribution]} \quad (5)$$

PMF factors dominated by a high proportion of secondary aerosol-related species (e.g., NO$_3^-$ or SOC) were selected as response variables, while the remaining factors were regarded as predictors. To explore the longer temporal variations in secondary aerosol contributions that are not captured in PMF modeling time intervals, the factor contributions were calculated using 24-h averages and applied in robust regression modeling.

3 RESULTS AND DISCUSSION

3.1 Data Pretreatment and Summary Statistics

To ensure the input data qualities were reliable, if the daily operation time of one of the online monitoring instruments (i.e., Xact 625i or IGAC S-611) was less than 67% (e.g., maintenance of
instruments), all measurements of that day were omitted. As a result, a total of 1344 hourly samples, which equaled 56 daily filter samples, were retained. During the selected sampling period, the air quality index (Tsai and Lin, 2021) showed that 70% of the values fell within the ‘good (≤ 50)’ category, while 26% were categorized as ‘moderate (51–100)’. The leading pollutant during the period was PM$_{2.5}$. As shown in Fig. 3, the major chemical compositions in PM$_{2.5}$ were secondary aerosol-related species, indicating the necessity to characterize the primary emission sources of these secondary aerosols to formulate effective control strategies for PM$_{2.5}$.

To enhance the applicability of 24-h resolution data in the PMF solution (Kuo et al., 2014), hourly online measurements were averaged into 3-h intervals since using 1-h and 2-h intervals would retrieved inadequate predicted values for organic components by the model. The scale difference of the temporal resolution was reduced from 1:24 to 1:8 between 3-h online measurements and 24-h filter samples. Wang et al. (2018) found that the PMF results were similar between using 3-h average PM$_{2.5}$ component data and those obtained from using hourly measurements.

In the first stage of PMF modeling, a six-factor solution was selected based on the IM and IS and the interpretability of the factor profiles. Species with low Spearman’s correlation coefficients (< 0.6) were removed from the model because of unsatisfactory predicted values. The species having high proportions (> 70%) of below detection limit and missing values were also excluded. By using the method described in Wu and Yu (2016), the estimated (OC/EC)$_{pri}$ value was 1.6 (Fig. S1), reasonably falling in the range reported by the literature (Li et al., 2016b). After data pretreatment, 26 species were selected as inputs. In the second stage, prior whole profile constraints were imposed on these factors.

### 3.2 Interpretation and Identification of Factor Profiles Obtained from Two-stage PMF

The concentrations and explained variations (EV), which represent the relative abundance of the species in retrieved profiles (Fig. 4), were utilized for factor profile interpretation (Anttila et al., 1995). Back trajectory analyses utilizing the HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model of the National Oceanic and Atmospheric Administration (NOAA) were implemented to assist in identifying the selected factors (Stein et al., 2015).

Factor 1 was characterized by high loadings of elemental K, Se, Br, and the organic tracer of DHOPA. K, Se, and Br may be associated with multiple sources, including incineration, coal burning,
Fig. 4. Factor profiles of each chemical species acquired from two-stage PMF approach (black bars denote concentration; gray points denote explained variation).

and combustions (Galvão et al., 2021; Han et al., 2006; Liu et al., 2005; Pokorná et al., 2016). K may also be related to traffic emissions and soil dust (Galvão et al., 2021). The presence of DHOPA revealed that this factor was potentially correlated with the emissions of toluene and xylenes, as they are typical markers of industrial solvents (Brown et al., 2007; Guo et al., 2004). This profile also showed abundant secondary species of NO$_3$– and SOC, which are the products converted via the oxidation of precursor gaseous pollutants, implying the involvement of contributions originating from primary feature sources. Hence, depending on the predominant characteristics of the secondary aerosols and their potentially diverse sources, Factor 1 was identified as aged emissions.

Factor 2 was loaded on multiple metals, such as Cr, Mn, Fe, Cu, Zn, and carbonaceous components of EC and POC. Fe, Cu, and Zn are commonly found in the tires and brake wears of vehicles...
through the coastline, indicating that the air pollutants emitted from port activities and surrounding highest contribution. The results revealed that the trajectory lines at a height of 500 m passed through the land for approximately 6 hours as estimated by HYSPLIT, and reached the coastline, which is situated several oil combustion-related sources, such as ship engines fueled by heavy oil, oil-based industrial boilers and electric power plants (Hsu et al., 2017; Laiman et al., 2022). The HYSPLIT model was applied on August 29 from 17:00 to 19:00, when Factor 4 had the highest contribution. The results revealed that the trajectory lines at a height of 500 m passed through the coastline, indicating that the air pollutants emitted from port activities and surrounding industrial sources were potentially transported by sea breeze to the study area (Fig. S2(A)). This could explain the moderate loading of Na⁺ and high 2-MT presented as well (Urban et al., 2012; Zhang et al., 2013). Factor 4 exhibited high loadings of V and Ni, which are distinctive markers of industrial oil combustion (Galvão et al., 2021; Karnae and John, 2011). The profile also constitutes a partial loading of Na⁺, implying that this factor was probably transported by the sea breeze from the coastline, which is situated several oil combustion-related sources, such as ship engines fueled by heavy oil, oil-based industrial boilers and electric power plants (Hsu et al., 2017; Laiman et al., 2022). The HYSPLIT model was applied on August 29 from 17:00 to 19:00, when Factor 4 had the highest contribution. The results revealed that the trajectory lines at a height of 500 m passed through the coastline, indicating that the air pollutants emitted from port activities and surrounding industrial sources were potentially transported by sea breeze to the study area (Fig. S2(A)). Factor 4 was identified as oil combustion.

Factor 5 was characterized by high loadings of levoglucosan and moderate loadings of Na⁺, EC, and POC. Levoglucosan is a distinguishing marker of biomass burning (Bhattarai et al., 2019; Feng et al., 2013), which is regarded as a combustion source and contributes EC and POC simultaneously (Singh et al., 2016; Titos et al., 2017). Comparing with species of K, levoglucosan can be more specific to biomass burning source since soil dust and seawater are significant contributors to potassium as well (Urban et al., 2012; Zhang et al., 2013). In addition, this factor was also characterized by abundant 2-MT, PA/iPA and a moderate loading of DHOPA in the profile. These organic tracers are chemically converted via VOCs precursors emitted from biogenic and anthropogenic sources. The mixing feature of this factor was investigated. The HYSPLIT model was applied to investigate the transport route of Factor 5 on August 21 from 20:00 to 22:00 when it showed the highest contribution. The results revealed that this factor was transported by the sea breeze from the coastline, then passed through the land for approximately 6 hours as estimated by HYSPLIT, and reached the sampling site (Fig. S2(B)). This could explain the moderate loading of Na⁺ and high 2-MT presented in the profile since the atmospheric isoprene derives from both terrestrial plants and marine phytoplankton (Li et al., 2018). The air pollution present along the regions in the path of the breeze was also transported, resulting in the mixed features of organic tracers. Because of the dominant presence of levoglucosan in the profile, Factor 5 was identified as biomass burning related.

Factor 6 exhibited high loadings of Ca and Ti, which are typically regarded as crustal materials and soil dust that can be transported by strong winds (Galvão et al., 2021; Watson et al., 2002). This factor showed a higher normalized contribution during hours 8 to 19 (average = 1.32) than hours 20 to 7 (= 0.68) in daily variation, consistent with the stronger wind speed in the daytime (average = 1.6 m s⁻¹) than nighttime (1.1 m s⁻¹). Given the dominant crustal species observed in the profile, Factor 6 was named soil dust.

Factor 7 served as the residual part of the modeling result of the 6-factor solution and was additionally differentiated via the model constraint in the second-stage PMF. The profile mainly loaded with NO₃⁻ with minor NH₄⁺, and thus was identified as nitrate-rich secondary aerosol. Factor 7 was not characterized by any input primary species, indicating non-association with retrieved primary air pollution sources in the 3-h temporal resolution, but may relate to the 24-h...
resolution since various factors affect the formation mechanism of NO$_3^-$ (Chuang et al., 2022). To further examine the origins of Factor 7, a robust regression using 24-h average contributions was performed to investigate the underlying associations between the retrieved factors.

### 3.3 Investigation of Associations between PMF Factors Using Robust Regression Model

Robust regression was conducted to characterize the associations between the PMF-retrieved factors and the origins of secondary aerosols. The factors of aged emissions and nitrate-rich secondary aerosol, which loaded with high proportions of secondary aerosol-related species, were individually selected as response variables and regressed against the remaining factors. The 3-h resolution factor contributions were calculated as 24-h averages in modeling to capture the latent correlations over a longer time span.

Overall, industry showed the highest contribution to aged emissions, whereas oil combustion was the largest contributor to nitrate-rich secondary aerosol (Fig. 5). These results indicated the importance of regulating specific factors to inhibit the formation of particular types of secondary aerosols. For example, oil combustion contributed significantly to nitrate-rich secondary aerosol but was low to aged emissions, implying that oil combustion may be more significant for the formation of NO$_3^-$. Furthermore, the results indicated that industry and vehicle related emissions were significant contributors to aged emissions (48% and 28%, respectively) and nitrate-rich secondary aerosol (30% and 22%, respectively), suggesting that the secondary feature factors, which mainly composed of SOC and NO$_3^-$, were largely formed from precursors emitted by industry and vehicles. Therefore, regulating both industry and vehicle related emissions could be often effective in inhibiting the formation of SOC and NO$_3^-$ over a longer period.

### 3.4 Factor Contributions to PM$_{2.5}$ and Secondary Aerosol-related Species

#### 3.4.1 Factor contributions obtained from two-stage PMF and robust regression models

This study used two models to allocate the contributions of PM$_{2.5}$, SIA species, and SOC to the primary feature factors. The first model was a two-stage PMF, which was used to characterize the primary and secondary feature factors and apportion preliminary contributions of the

![Graph showing average contributions of each primary feature factors to secondary feature factors obtained from robust regression model.](https://example.com/graph.png)
aforementioned pollutants using 3-h measurements. The second model was a robust regression model that utilized 24-h average inputs to re-apportion the contributions of secondary feature factors to primary feature factors based on the predictor coefficients derived from the model. Finally, contributions of PM$_{2.5}$, SIA species, and SOC retrieved from both models were merged by each primary feature factor.

For PM$_{2.5}$ (Fig. 6(A)), the largest PM$_{2.5}$ contributor was oil combustion (contribution = 3.66 µg m$^{-3}$), and followed by industry (3.56 µg m$^{-3}$), vehicle related (3.09 µg m$^{-3}$), biomass burning related (2.23 µg m$^{-3}$), and soil dust (1.85 µg m$^{-3}$). For vehicle related, the significant difference between PM$_{2.5}$ and the sum of SIA species and SOC (1.05 µg m$^{-3}$) implied that the majority of PM$_{2.5}$ of this factor were dominantly constituted with primary species, including POC, EC or Fe. In addition, the SIA species of SO$_4^{2-}$ (contribution = 1.48 µg m$^{-3}$), NO$_3^-$ (0.76 µg m$^{-3}$) and NH$_4^+$ (1.11 µg m$^{-3}$) were mainly contributed by oil combustion, suggesting that the control of oil combustion could be more cost-effective for reducing the formation of SIA species than other factors. Numerous studies have pointed out that shipping emissions of SO$_x$ and NO$_x$, which are crucial precursors of SO$_4^{2-}$ and NO$_3^-$, affect the ambient air quality considerably due to the combustion of heavy oil with high sulfur content, resulting in high emissions amount of these pollutants (Chow et al., 2010; Ni et al., 2020; Nunes et al., 2017; Viana et al., 2014; Xiao et al., 2018). Regarding SOC, the largest contributor was industry (0.43 µg m$^{-3}$), highlighting the need for prioritizing its management.

A higher contribution percentage of soil dust (13%) to PM$_{2.5}$ mass was retrieved than in previous studies (Heo et al., 2009; Liao et al., 2017; Liu et al., 2006). This could be attributed to the location of the sampling site in a redevelopment zone surrounded by several buildings under construction, leading to a mixture of anthropogenic emissions from construction activities, which are potential sources of Ca and Ti (Galvão et al., 2021).

In PMF results (Fig. 6(B)), oil combustion showed the greatest contribution to PM$_{2.5}$ (2.82 µg m$^{-3}$), followed by biomass burning related (1.92 µg m$^{-3}$) and vehicle related (1.71 µg m$^{-3}$). Furthermore,
oil combustion also presented the highest sum of SIA species (2.67 μg m⁻³), including the largest contribution to SO₄²⁻ (1.37 μg m⁻³), NO₃⁻ (0.37 μg m⁻³) and NH₄⁺ (0.93 μg m⁻³). Since PMF was modeled using the measurements with a 3-h resolution, suggesting that the SIA species were mainly formed through the emission of oil combustion and transported to the receptor site over this temporal period. On the other hand, SOC (0.24 μg m⁻³) was mostly contributed by biomass burning related. In the results of robust regression using 24-h average inputs (Fig. 6(C)), the majority of secondary aerosol-related species were distributed in industry (sum = 1.65 μg m⁻³), pointing out a longer duration for the formation of secondary aerosols associated with the emission of industry.

To facilitate the formulation of air pollution control strategies, the contributions derived from PMF and the robust regression model were further interpreted based on different apportionment principles and temporal settings, with 3-h and 24-h resolutions as inputs for PMF and robust regression, respectively. The SO₄²⁻ contribution from industry was selected as an example for illustration. The 0.24 μg m⁻³ apportioned by PMF and 0.31 μg m⁻³ attributed by the robust regression model together showed that industry accounted for a total contribution of 0.55 μg m⁻³ in SO₄²⁻. The 0.24 μg m⁻³ represents the SO₄²⁻ directly and indirectly (e.g., interact with meteorological factors) formed by industry emissions in three hours. The robust regression apportioned 0.31 μg m⁻³ denotes the SO₄²⁻ in secondary feature factors (i.e., aged emissions and nitrate-rich secondary aerosol) formed by industry emissions in 24 hours. The implication is that if the emission of industry can be properly controlled in three hours, the additional 0.31 μg m⁻³ of SO₄²⁻ contribution by industry is expected to be reduced as inhibiting the formation of SO₄²⁻ in the secondary feature factors. A sensitivity analysis was conducted to evaluate temporal resolution issues. The 3-h resolution measurements were averaged over 6-hour intervals and applied to two-stage PMF modeling. Similar factor profiles as shown in Fig. 4 were obtained. Next, the retrieved PMF factors were calculated as 24-h averages for robust regression modeling to attribute the contributions. Results showed that industry contributed a total of 0.57 μg m⁻³ of SO₄²⁻, whereas 0.31 μg m⁻³ was apportioned by PMF and 0.26 μg m⁻³ was allocated by the robust regression model. A similar total contribution was obtained using the 6-h and 3-h measurements. However, an increased contribution was apportioned by PMF with 6-h measurements, and a decreased contribution was retrieved by the regression analysis. This is likely because the extended period provided a longer reaction time for the formation of secondary aerosols. The sensitivity analysis results suggested that implementing controls on air pollution sources at an early stage can effectively inhibit the formation of secondary aerosols with significant contributions.

### 3.4.2 Estimated SOC contributions between PMF and SOA-tracer method

In this study, three organic molecular tracers of SOA, namely 2-MT, DHOPA and PA/iPA were applied to PMF for source characterization and SOC apportionment. To assess the applicability of using organic tracers, the SOA-tracer method was used and the results were compared with the PMF modeling results. The SOA-tracer method utilizes the SOC mass fraction determined from the smog chamber experiment and tracer concentration to estimate SOC (Ding et al., 2014; Kleindienst et al., 2007; Lanzafame et al., 2021; Rutter et al., 2014; Srivastava et al., 2021). For the detailed information on the SOA-tracer method, please refer to Text S1 in Supplementary material. The tracer concentrations in each factor profile were utilized to estimate the SOC with the SOA-tracer method. The resulting values were summed (SOC_{SOA-tracer}) and compared with the SOC apportioned by PMF (SOC_{PMF}).

The top three SOC contributors estimated by two methods were the same, which were aged emissions (SOC_{SOA-tracer} = 0.47 μg m⁻³ vs. SOC_{PMF} = 0.70 μg m⁻³), biomass burning related (0.31 vs. 0.24 μg m⁻³) and industry (0.06 vs. 0.10 μg m⁻³), although differences were found in the quantified contributions. Incorporating more critical organic tracers (e.g., cholesterol as a marker species of cooking) into PMF analysis may help improve the coherence of estimated SOC between two methods. Overall, a good agreement ($R^2 = 0.74$) was found between the total SOC_{SOA-tracer} and SOC_{PMF} (Fig. S3), indicating the satisfactory consistency as found in previous studies (Lanzafame et al., 2021; Zhang et al., 2009). In both methods, the primary feature factor of industry exhibited the highest SOC contribution, which could be responsible for DHOPA. Based on the formula of SOA-tracer method (Text S1), DHOPA showed a high measured tracer concentration but a low
3.5 Comparison with Results Using One-stage PMF Modeling

The conventional one-stage PMF modeling using S/N categories as ‘strong’ for SIA species and SOC with 6- and 7-factor solutions were conducted and compared with the developed two-stage PMF. The factor profiles were identified and illustrated in Fig. S4 (6-factor solution) and SS (7-factor solution) of the Supplementary material.

In the 6-factor solution, an independent factor of secondary inorganic aerosol with a high EV value of 81% NO$_3^-$ in the profile was retrieved. The concentrated SIA species in one factor made it difficult to recognize the related primary sources. Conversely, the independent factor of secondary inorganic aerosol was not identified in the first-stage PMF of our approach by using ‘weak’ settings. This suggested that the target for apportionment (such as SIA species and SOC in this study) should be set as ‘weak’. Furthermore, the oil combustion and biomass burning related were mixed into a factor, possibly because the number of factors was set at six, which prompted a 7-factor solution. The 7-factor solution also yielded a high EV value of 91% NO$_3^-$ in the profile (Fig. S5). In contrast, species of NO$_3^-$ and NH$_4^+$ were distributed in most of the profiles retrieved from the two-stage PMF (Fig. 4). In addition, physically unreasonable results were obtained from the conventional 7-factor solution, such as 0% contribution of biomass burning related to NO$_3^-$ (Fig. S5). In previous studies, high levels of NO$_3^-$ were typically observed during biomass burning influence periods (Behera and Balasubramanian, 2014; Cheng et al., 2009; Ezcurra et al., 2001), implying that NO$_3^-$ should be loaded into the profile of biomass burning related. The uninterpretable results from conventional PMF demonstrated that the two-stage PMF approach developed in this study was better suited for apportioning secondary aerosol-related species.

Chen et al. (2022) proposed an alternative approach to quantify the contributions of inorganic aerosols (IA, the sum of SO$_4^{2-}$, NO$_3^-$, and NH$_4^+$ concentrations) using PMF and a multiple linear regression model. In PMF modeling, only metals were considered. Hence, no secondary feature factors were retrieved. Next, the PMF factors were applied as predictors whereas the IA was introduced as a response variable in the regression analysis. Consequently, the contribution of IA to each factor was estimated. In contrast, this study input both primary and SIA species simultaneously into the PMF analysis, which offers several advantages. First, the resulting profiles consisted of both primary and secondary species, better reflecting the real-world situation in which these species may coexist. For example, in the profile of aged emissions, the presence of K, Se, Br, and DHOPA suggests that the formation of SOC and NO$_3^-$ may relate to industrial sources. This finding was further supported by the robust regression results, which indicated that 48% of the aged emissions were attributed to industry. Second, finer quantitative contributions were revealed. As described in Section 3.4.1, PMF apportioned 0.24 µg m$^{-3}$ of SO$_4^{2-}$ contribution to industry, while robust regression allocated 0.31 µg m$^{-3}$. The separation of contributions by the two methods can provide a comprehensive understanding of the formation of secondary aerosols, highlighting the importance of incorporating SIA species into PMF analysis. The fusion strategy developed in this study can be utilized to formulate a more refined control strategy for secondary aerosols, which will be beneficial to policymakers.

In summary, to address the individual secondary feature factors derived from conventional one-stage PMF (Liao et al., 2017; Oduber et al., 2021; Park et al., 2019; Sharma et al., 2016; Taghvaei et al., 2018), the proposed fusion strategy of two-stage PMF and robust regression modeling can effectively apportion secondary aerosols to primary emission sources. Moreover, our approach delivered a refined result, which approximated real-world situations and allowed for exploring various temporal distributions of secondary aerosols formed from primary sources, comparing to the approach in Chen et al. (2022). These findings highlight the advantages and uniqueness of the proposed fusion model.

3.6 Study Limitations

In this study, PMF and robust regression were fused to provide refined apportionment results on the temporal scale, aiding in the development of air pollution control strategies related to the
formation of SIA species and SOC, although several limitations can be further improved. First, as in most PMF studies, source regions cannot be identified with receptor modeling alone. To address this limitation, the transport route of biomass burning related was characterized through back trajectory analysis using the HYSPLIT model. However, to recognize the source region more accurately, detailed spatiotemporal information (e.g., satellite data of active fire products and vegetation index) could be introduced to facilitate the identification of possible locations of agricultural burning. Moreover, integrating local wind data and trajectory analysis can assist in determining the source direction and distinguishing local and regional sources, and provide insights into the chemical transformations over various time scales as well as transportation processes between known sources and the receptor. Second, four organic molecular tracers were applied to characterize the PM$_{2.5}$ factors in this study. For instance, levoglucosan was applied and the factor of biomass burning related was identified. For future studies, the inclusion of more organic tracers is suggested for retrieving the refined modeling results and improving factor identifications. Previous studies have shown that incorporating the species of cholesterol can be beneficial for the sampling site surrounded by large residential and commercial areas to distinguish the cooking source (Al-Naiema et al., 2018; Yu et al., 2021). Additionally, the associated precursors (e.g., toluene and naphthalene, which are regarded as hazardous air pollutants) of these organic tracers are also recommended to be included in the analysis in future studies to comprehensively understand the formation of secondary aerosols and recognize potential risk concerns of pollution sources (Liao et al., 2015; U.S. EPA, 2023; Wu et al., 2007).

4 CONCLUSIONS

In this study, we developed a fusion strategy combining two-stage PMF and robust regression modeling to better understand the contributions of PM$_{2.5}$ secondary aerosols originating from primary air pollution sources. Through this approach, a factor dominated by nitrate was additionally retrieved and the secondary feature factors were re-apportioned. The combined use of these two models characterized the origins of secondary aerosols and provided refined results in different temporal scales. That is, emission and generation within three hours, accompanied by the further formation in 24 hours. Our findings provide valuable insights for formulating effective strategies to control air pollution.

ACKNOWLEDGMENTS

This study was financially supported by National Science and Technology Council of Taiwan (grant numbers 109-2221-E-002-057-MY3 and 110-2811-E-002-502-MY3) and Department of Environmental Protection of Taipei City Government (111S058). This article was subsidized for English editing by National Taiwan University under the Excellence Improvement Program for Doctoral Students (grant number 108-2926-I-002-002-MY4), sponsored by National Science and Technology Council, Taiwan.

SUPPLEMENTARY MATERIAL

Supplementary material for this article can be found in the online version at https://doi.org/10.4209/aaqr.230121

REFERENCES


from aromatic hydrocarbons. Atmos. Chem. Phys. 17, 2053–2065. https://doi.org/10.5194/acp-17-2053-2017


Jain, S., Sharma, S.K., Vijayan, N., Mandal, T.K. (2020). Seasonal characteristics of aerosols (PM$_{2.5}$ and PM$_{10}$) and their source apportionment using PMF: A four year study over Delhi, India. Environ. Pollut. 262, 114337. https://doi.org/10.1016/j.envpol.2020.114337


Kleindienst, T.E., Jaoui, M., Lewandowski, M., Offenberg, J.H., Lewis, C.W., Bhave, P.V., Edney, E.O.


Tsai, W.T., Lin, Y.Q. (2021). Trend analysis of air quality index (AQI) and greenhouse gas (GHG) emissions in Taiwan and their regulatory countermeasures. Environments 8, 29. https://doi.org/10.3390/environments8040029

carbon and PM$_{2.5}$: a case study in the Pearl River Delta, China. ACS Earth Space Chem. A 1, 562–571. https://doi.org/10.1021/acsearthspacechem.7b00088


