Analysis of non-linear relationship of PM2.5 mass concentration with aerosol extinction coefficient and RH in Hefei, China

Yinan Chen1,2,3, Shiguo Zhang1,2,3,4, Yonghua Wu5, Kee Yuan1,3, Jian Huang1,3, Dongfeng Shi1,3, Shunxing Hu1,3*

1Key Laboratory of Atmospheric Optics, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China
2Science Island Branch of Graduate School, University of Science and Technology of China, Hefei 230026, China
3Advanced Laser Technology Laboratory of Anhui Province, Hefei 230037, China
4Anhui Meteorological Observation Technical Center, Hefei 230031, China
5Optical Remote Sensing Lab, the City College of New York (CCNY), New York, NY, 10031, USA

* Corresponding author: Shunxing Hu (sxhu@aiofm.ac.cn)

Abstract

Optical remote sensing of PM2.5 concentration complements ground-based in situ sampler observations and depends on aerosol extinction-to-mass conversion and meteorological factors. Based on Mie scattering theory, we derived a non-linear analytical expression among the PM2.5 mass concentration, aerosol extinction coefficient, and hygroscopic factor. We analyzed 1-year data of aerosol size distribution, PM2.5, visibility, and relative humidity (RH) and refined the exponential factors for improving PM2.5 estimated from the aerosol extinction coefficient in Hefei, China. Aerosol size distribution data on adjacent days were used to approximate the hygroscopic factors of the aerosols. The measured PM2.5 in 2020 was used to refine the exponential factors relevant to the large aerosol extinction coefficient and RH in the analytical function. Next, we evaluated the
derived PM$_{2.5}$ from aerosol extinction with the tapered element oscillating microbalance (TEOM), which independently observed PM$_{2.5}$ in January 2021; their comparisons indicated good consistency and strong correlation with a linear correlation coefficient of R$^2$=0.76. Finally, by applying the analytical function of PM$_{2.5}$ and aerosol extinction, we obtained the spatial and vertical distribution of PM$_{2.5}$ from the NASA/CALIPSO-observed aerosol extinction profiles, which showed reasonable agreement and good correlation with the TEOM-measured PM$_{2.5}$ in several major cities in Anhui Province, China.

*Keywords:* PM$_{2.5}$, Aerosol size distribution, Extinction coefficient, Hygroscopic factor
1 INTRODUCTION

Atmospheric aerosols are mixtures of solid and liquid particles suspended in the atmosphere that affect air quality, climate, and human health (Kampa and Castanas, 2008). Particulate matter (PM) with an aerodynamic diameter of ≤2.5 μm is known as PM$_{2.5}$; it is one of the major air pollutants that affects human health. Studies have demonstrated that increased PM$_{2.5}$ concentration is associated with morbidity and mortality from multiple diseases, particularly in individuals with cardiovascular and respiratory diseases (Brauer et al., 2016; Xue et al., 2022). Long-term exposure to high levels of PM$_{2.5}$, directly or indirectly, significantly affects health. Therefore, monitoring ground PM$_{2.5}$ mass concentration is necessary for the development of air pollution prevention and control measures.

PM$_{2.5}$ mass concentrations are typically measured at ground-based air quality or environmental monitoring sites. However, such sites are limited in number and spatially uneven, limiting the knowledge of spatial variations in PM$_{2.5}$. The remote sensing measurements from satellites and lidar sensors provide a comprehensive aerosol spatial distribution with a global-regional coverage, complementing limited in situ measurements at ground sites. Research on remote sensing applications for air quality monitoring has made significant progress (Hoff and Christopher, 2009; Zhang et al., 2021). Studies have used data from active and passive remote sensing observations, such as from an aerosol optical depth (AOD) product from the Moderate Resolution Imaging Spectroradiometer (Ma et al., 2014; Zeydan and Wang, 2019), the Visible Infrared Imaging Radiometer Suite (Wu et al., 2016; Yao et al., 2019), and the Cloud-Aerosol Lidar
with Orthogonal Polarization (CALIOP) (Sun et al., 2021; Zhang et al., 2019). However, remote sensing techniques do not directly measure PM$_{2.5}$ mass concentration; they rely on the correlation between aerosol optical parameters (aerosol extinction and optical depth) and PM$_{2.5}$ concentration. Various statistical methods have been applied in PM$_{2.5}$ estimates using satellite-retrieved AOD, including semi-empirical physical methods (Li et al., 2016; Tian and Chen, 2010), multiple linear and non-linear regression (Kloog et al., 2012; Liu et al., 2009b), geo-weighted regression (Ma et al., 2014; Yang et al., 2019), machine learning, and neural networks (Danesh Yazdi et al., 2020; Gündoğdu et al., 2022). These research methods have yielded relatively accurate, high-resolution PM$_{2.5}$ mass concentrations, thus effectively compensating for the insufficient data from ground monitoring stations. Although significant progress has been made with the aforementioned statistical techniques, physically interpreting the results is challenging because of the reliance on vast volumes of datasets, statistical models, and ambiguous connections between parameters.

The relationship among aerosol mass, extinction, and relative humidity (RH) is of great interest in interpreting the AOD–PM$_{2.5}$ correlation. Pilat and Ensor (1971) analyzed the aerosol mass extinction factor and showed a varied magnitude relevant to aerosol emission sources under different atmospheric conditions. According to Mie scattering theory, the relationship between the mass concentration of PM and the extinction coefficient is consistent with a positive correlation (Tao et al., 2016). Because of the influence of the main physical characteristics of atmospheric aerosols (size distribution, complex refractive index, and aerosol mass density), the aforementioned method developed to convert extinction coefficients to mass concentrations has a solid theoretical foundation.
Studies in China have shown that RH affects the optical properties of aerosols and mass concentrations via hygroscopic growth and chemical reactions under severe air pollution scenarios (Cao et al., 2012; Cheng et al., 2017; Lin et al., 2013; Qu et al., 2015; Tao et al., 2021). Relevant studies have shown that high RH may strengthen the hygroscopic growth and oxidization efficiency of aerosol precursors, thus affecting aerosol extinction, chemical components (e.g., sulfate and nitrate), and mass extinction efficiency (Cao et al., 2012; Cheng et al., 2017; Qu et al., 2015).

Particle size distribution also varies regularly with increasing RH, mass fraction of hydrophilic chemical components in PM$_{2.5}$, and their mass median aerodynamic diameters, becoming larger for ambient RH > 60% than those for RH < 60% (Tao et al., 2021). Although the optical parameters calculated using the particle number density distribution are meaningful, some relevant parameters must be considered (Lin et al., 2013).

To convert aerosol extinction to PM$_{2.5}$, most studies have used linear models by fitting aerosol extinction coefficients to simultaneously measured PM$_{2.5}$ mass concentrations (Lyu et al., 2018; Tao et al., 2016; Xiang et al., 2020) or have assumed empirical parameters (Ma et al., 2020; Toth et al., 2019; Toth et al., 2022). The obtained equation can be used to convert the extinction coefficient profile into a mass concentration profile by assuming constant aerosol characteristics or types. Taking aerosol hygroscopicity into account, some researchers have employed empirical hygroscopic factor formulas for humidity correction on the assumed aerosol type (Ma et al., 2020; Toth et al., 2019); however, other studies have ignored the change in hygroscopic factor when the RH was low (<60–70%) (Lyu et al., 2018; Tao et al., 2016; Xiang et al., 2020). However, these studies did not conduct simultaneous observations of aerosol size distribution, PM$_{2.5}$, visibility or
extinction, and RH at the same site. Importantly, the analytical expression of PM$_{2.5}$, aerosol extinction, and hygroscopic factors may vary significantly with aerosol type and size distribution, particularly under high RH and severe air pollution conditions.

This study aimed to derive a novel non-linear analytical expression for PM$_{2.5}$, aerosol extinction, and hygroscopic factors. Co-located observations of the aerosol size distribution, PM$_{2.5}$, visibility, and RH were used to refine the exponential and hygroscopic factors in the analytical function. We further evaluated the derived PM$_{2.5}$ from aerosol extinction by using a tapered element oscillating microbalance (TEOM), which independently measured PM$_{2.5}$ concentration from a national environmental monitoring site in Hefei, China. Regional and vertical distributions of PM$_{2.5}$ were further derived from the NASA space-borne lidar CALIPSO/CALIOP aerosol extinction product in Anhui Province, China. The paper is organized as follows. Section 2 describes the methodology used in the study, Section 3 introduces the instruments used and the data collection procedure, Section 4 provides the results and discussion, and Section 5 presents the conclusions.

2 METHODOLOGY

According to Mie theory, the aerosol extinction coefficient in ambient air ($\sigma_{RH}$) is expressed as follows (Koelemeijer et al., 2006, Hoff et al., 2009):

$$
\sigma_{RH} = \int_{0}^{\infty} \pi Q_{ext}(m, r, \lambda) r^2 n(r)_{RH} dr \approx \pi \bar{Q}_{ext} \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr
$$

(1)
where $r$ denotes the particle radius (unit: $\mu m$), $m$ represents the complex refractive index, and $Q_{ext}$ represents the particle extinction efficiency factor for a single particle at $m$, $r$, and $\lambda$. $n(r)_{RH} = dN/dlgr$ denotes the particle size distribution under RH (unit: cm$^{-3}$). $\bar{Q}_{ext}$ is the overall average extinction efficiency factor (Koelemeijer et al., 2006, Hoff et al., 2009).

Dry PM$_{2.5}$ mass concentration is expressed as follows:

$$C_{PM2.5} = \frac{1}{3} \pi \rho r^3 n(r)_{dry} dr \cong \frac{4\pi\bar{\rho}}{3} \int_{r_{min}}^{r_{max}} r^3 n(r)_{dry} dr$$  \hspace{1cm} (2)$$

where $r$ denotes the particle radius (unit: $\mu m$), $\rho$ is the dry aerosol mass density (unit: g cm$^{-3}$), $n(r)_{dry}$ denotes the dry particle size distribution (unit: cm$^{-3}$), and $\bar{\rho}$ represents the mean effective density (Koelemeijer et al., 2006, Hoff et al., 2009).

Taking the logarithm of both sides of Eq. (1) and multiplying by a constant $a = \log(\pi\bar{Q}_{ext})\left(\frac{4\pi\bar{\rho}}{3}\right)$, Eq. (1) becomes:

$$a \ln \sigma_{RH} = a \ln [\pi\bar{Q}_{ext}] + a \ln \left|\int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr\right|$$  \hspace{1cm} (3)$$

Then, for Eq. (2), taking the logarithm of both sides, we obtain

$$\ln C_{PM2.5} = \ln \frac{4\pi\bar{\rho}}{3} + \ln \int_{r_{min}}^{r_{max}} r^3 n(r)_{dry} dr$$  \hspace{1cm} (4)$$

We subtract Eq. (3) from Eq. (4) (the left and right sides of the two equations are subtracted,
respectively), and the results are as follows:

\[
\ln C_{PM2.5} - a(\ln \sigma_{RH}) = \ln \left( \frac{4\pi \bar{\rho}}{3} \right) - a(\ln |\pi \bar{q}_{\text{ext}}|) + \ln \int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r)_{\text{dry}} dr - a \left( \ln \left[ \int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{RH} dr \right] \right)
\]  

(5)

The value of “a” is \( \log(\pi \bar{q}_{\text{ext}}) \left( \frac{4\pi \bar{\rho}}{3} \right) \), thus \( \ln \left( \frac{4\pi \bar{\rho}}{3} \right) - a(\ln |\pi \bar{q}_{\text{ext}}|) = 0 \), then Eq. (5) becomes:

\[
\ln C_{PM2.5} - a(\ln \sigma_{RH}) = \ln \int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r)_{\text{dry}} dr - a \left( \ln \left[ \int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{RH} dr \right] \right)
\]  

(6)

Furthermore, by moving \( a(\ln \sigma_{RH}) \) to the right side of Eq.(6), setting \( b = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r)_{\text{dry}} dr}{\left( \int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{RH} dr \right)^{\frac{3}{2}}} \), and then removing the logarithm on both sides from Eq. (6), we obtained the analytical relationship between the PM2.5 mass concentration and aerosol extinction(\( \sigma_{RH} \)) by using Eq. (7).

\[
C_{PM2.5} = b \sigma_{RH}^a
\]  

(7)

According to the literature (Junge, 1955; Tao et al., 2021), particle size increases significantly owing to aerosol hygroscopic growth, and particle size distribution varies regularly with increasing RH. Thus, the change in the aerosol number size distribution associated with the dry PM2.5 mass concentration must be considered.
\[
b = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r) \, dr}{(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r) \, dr)^{\frac{a}{2}}} = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r) \, dr}{(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r) \, dr)^{\frac{a}{2}}} \left( \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r) \, dr}{(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r) \, dr)^{\frac{a}{2}}} \right)^a
\]

where \( b_{\text{dry}} = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^3 n(r) \, dr}{(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r) \, dr)^{\frac{a}{2}}} \); \( f(RH) \) is a function associated with RH and is called the power of the hygroscopic factor of the aerosol surface area. \( f(RH) \) is simply denoted as the hygroscopic factor in the following discussions. Then, an analytical expression between PM\(_{2.5}\) mass concentration and the aerosol extinction coefficient can be derived as follows:

\[
C_{\text{PM2.5}} = \frac{b_{\text{dry}} \sigma \rho_{RH}^{a}}{f(RH)}
\]

where \( a \) is a constant, the unit of \( b \) is \( \frac{\mu g}{m^3} \cdot \left( \frac{1}{m} \right)^a \), the unit of \( \sigma_{RH} \) is \( m^{-1} \), and the unit of \( C_{\text{PM2.5}} \) is \( \mu g \cdot m^{-3} \). Further details on the derivation in Section 1 (Methodology) are provided in the supplementary material.

### 3 EXPERIMENTAL INSTRUMENTS AND DATA COLLECTION

Data on PM\(_{2.5}\) mass concentration, aerosol size distribution, visibility, and RH were collected from a field campaign conducted at the Hefei Meteorological Science and Technology Park (117.2388° E, 31.8652° N) from January 1 to December 31, 2020. In addition, PM\(_{2.5}\) was directly...
measured using a TEOM at the national standard air quality monitoring station (labeled as 1274A, 1.3 km from our study site) operated by Anhui Environmental Protection Agency, and these data were used to validate our PM$_{2.5}$ estimate.

At our study site, the PM$_{2.5}$ mass concentration was measured using a PM sensor (SPS30, Sensirion, Switzerland). The SPS30 measurement principle is based on laser scattering with a 0.3 µm lower detection limit. A laser beam strikes a particle and causes it to scatter. The sensor detects the intensity and angle of the beam, and the algorithm determines the concentration. The accuracy of measurements of the mass concentration of PM$_{2.5}$ (0–100 µg m$^{-3}$) is ± 10 µg m$^{-3}$, and the accuracy of measurements from 100 to 1000 µg m$^{-3}$ is ± 10% (Kirešová et al., 2022). Studies have shown that SPS30 has a limited measurement range (Kuula et al., 2020). If the measured size fraction is within the sensor's limited detection range, the sensor may provide accurate results, which is nearly of the regulatory grade. Although SPS30 is a low-cost optical PM sensor, field and laboratory evaluations indicate that its measurements of PM$_{2.5}$ show a strong correlation with the TEOM-reported PM$_{2.5}$ ($R^2 = 0.80$) and track the PM$_{2.5}$ diurnal variations efficiently (Kuula et al., 2020; South Coast Air Quality Management District., 2019).

RH data were obtained using the Humidity and Temperature Probe HMP155A with the HUMICAP®180R sensor developed by Vaisala. RH was measured with an accuracy of ± (1.0% ± 0.008 × reading) % at temperatures between -20°C and 40°C.

The atmospheric visibility (VIS) was measured using a forward scattering visibility meter (HY-V35) over a measurement range of 10–35 km. HY-V35 was provided by Huayun Sounding Meteorological Technology Corporation (Beijing, China). The forward scattering visibility meter
estimates the optical range by measuring the scattering coefficient of a light pulse (Du et al., 2021). The primary sources of error in visibility measurements are scattering phase function changes caused by aerosol properties and multiple scattering under severe air pollution or fog conditions (Du et al., 2021). According to the findings of (Li and Sun, 2009), given the current state of technology, the relative uncertainty of visibility measurements is approximately 4%.

The Optical Particle Counter (OPC-N3) was developed by Alphasense Ltd. for the real-time monitoring of aerosol particle size distribution. The device uses diode laser light (wavelength 658 nm) and an elliptical mirror that reflects particle scattering into a detector (Nurowska et al., 2022). It measures airborne particle size distribution from 0.35 to 40 µm within 24 channels per minute at a flow rate of 5.5 L min⁻¹. The observation period was from October 9 to 13, 2020.

The relationship between the VIS and the extinction coefficient can be expressed as follows (WMO, 1996):

\[ \sigma_{RH} = \sigma_{ext} - \sigma_g = \frac{-\ln(\varepsilon)}{VIS} - \sigma_g, \]  

where \( \varepsilon \) denotes the contrast threshold for identifying targets, the value is 0.05; \( \sigma_{ext} \) represents the atmosphere extinction coefficient; \( \sigma_{RH} \) represents the aerosol extinction coefficient; and \( \sigma_g \) represents the molecular extinction coefficient.

The NASA space-borne lidar CALIOP instrument on board the CALIPSO satellite platform observes the global aerosol/cloud vertical distribution and provides aerosol type classification and optical property products (Winker et al., 2009). In this study, the vertical distribution of the level
2 aerosol extinction coefficient at 532 nm (Version-4.20 5-km Aerosol Profile) was used when CALIPSO passed over our study area. The vertical distribution of RH data is sourced from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) data which is a product of the Data Assimilation System of the Global Modeling and Assimilation Office (GMAO).

Fog, rainfall, and snowfall data were excluded to avoid the effects of complex meteorological conditions. In addition, data that reach the measurement limits of the instrument must be excluded because their accuracy is degraded. All experimental data were processed using the hourly mean data. In total, 4,844 groups of effective PM$_{2.5}$ mass concentration, RH, and visibility data and 120 groups of effective aerosol particle size distribution data were obtained. We excluded TEOM data where the PM$_{10}$ mass concentration was lower than that of PM$_{2.5}$. Finally, 4,172 groups of PM$_{2.5}$ mass concentrations from TEOM were used.

4 RESULTS AND DISCUSSION

4.1 Parameter Calculation Of PM$_{2.5}$ Mass Concentration Expressions: $a$, $b_{dry}$, $f(RH)$

A total of 643 sets of hourly averaged aerosol size distributions from the observed data were analyzed. We selected the OPC data on the adjacent sunny days (October 9–13) with a northeasterly light breeze and temperatures ranging from 14 °C to 24 °C. A reasonable assumption is that the changes in atmospheric RH play a major role in the variations in aerosol spectrum distribution. Thus, using the OPC data, we can approximate the hygroscopic growth factor $f(RH)$ in Eqs. (8)
and (9). Because some uncertainties may arise due to the aforementioned assumptions, \( f(RH) \) was further refined or optimized by minimizing the differences between the estimated and measured PM\(_{2.5} \) concentration.

As mentioned in the literature (Lyu et al., 2018; Tao et al., 2016), the aerosol hygroscopic growth can be neglected when the ambient RH is <60% and the nephelometer measurements are considered as dry scattering coefficients when the RH <60% (Cao et al., 2012). Tao et al. (2021) indicated that particle size distribution varies regularly with increasing RH, with particle diameter being larger at RH >60% compared to RH <60%. As a result, we refer to the “dry aerosol size distribution” as the average size distribution for RH <60%. The aerosol volume size distribution under various RH conditions is shown in Fig. 1(a). The distribution of the particle volume with a radius \(<1 \mu m \) was highly correlated with the RH, the volume density of particles \(<1 \mu m \) increases with increasing RH. Moreover, we calculated the results of the fine-mode (\( r \leq 1.25 \mu m \)) volume ratio at RH \( \geq 60\% \) to those at RH \(< 60\% \) are shown in Table S1. The volume ratio increases as the RH rises, which generally reflects the hygroscopic properties of aerosols.

Based on the statistical analysis of the physical and chemical properties of atmospheric aerosols and the actual measurement data in the Hefei area (Hu et al., 2006; Hu et al., 2019; Li et al., 2007; Ma et al., 2020), other relevant parameters are taken as \( m = 1.5 - 0.01i \) and \( \bar{\rho} = 2 \text{ g cm}^{-3} \) at \( \lambda = 0.55 \mu m \).

From the average dry aerosol size distribution in Fig. 1(a), the \( Q_{\text{ext}} = 2.9 \) was obtained. Substituting Eq. (8) and Eq. (5), we obtained \( b_{\text{dry}} = 0.195 \) and \( a = 0.96 \).

The hygroscopic factor \( f(RH) \), defined in Eq. (8), was calculated using the size distribution.
for RH >60%. As shown in Fig. 1(b), $f(RH)$ increases with the RH. By fitting the relationship between $f(RH)$ and RH using the power function yield, we obtained the following equation:

$$f(RH) = 3.302 \left( \frac{RH}{100} \right)^{7.211} + 1$$  \hspace{1cm} (11)

The hygroscopic factor has geographical characteristics and is related to aerosol sources. In the studies of Liu and Chen (Chen et al., 2014; Liu et al., 2009a), $f(RH = 80\%)$ was $1.63 \pm 0.19$ (in Beijing, China) and $1.58 \pm 0.22$ (in Tianjin, China), respectively, which is similar to the result of $f(RH = 80\%) = 1.66$ fitted in this study. The similarity may be because the study areas are inland cities with similar aerosol properties.

Applying Eq. (11) and the calculated values of $a$ and $b_{dry}$ in Eq. (9), the PM$_{2.5}$ mass concentration is shown in Eq. (12):

$$C_{PM2.5} = \frac{0.195}{3.302 \left( \frac{RH}{100} \right)^{7.211} + 1} \cdot (1000 \cdot \sigma_{RH})^{0.96},$$  \hspace{1cm} (12)

where the extinction coefficient is in units of km$^{-1}$ and $C_{PM2.5}$ is $\mu$g m$^{-3}$.

### 4.2 Correction For The Parameters $a$ And $f(RH)$

Previous studies have shown that the measurement range of SPS30 is limited (Kirešová et al., 2022) because the SPS30 measurement principle is different from that of the national standard measurement method (GB/T 15432-1995, Gravimetric method). In order to assure the data quality of SPS30, the ground-measured hourly average PM$_{2.5}$ mass concentrations from the SPS30 sensor...
were compared with data from the TEOM sampler at national standard air quality monitoring station 1274A. A comparison of the results is presented in Fig. S1. They showed a good linear correlation, with a linear correlation coefficient $R^2$ of 0.75 and a mean absolute error (MAE) of 9.70 $\mu$g m$^{-3}$. Overall, the SPS30 PM sensor provided reliable results.

We used Eq. (12) to calculate the PM$_{2.5}$ mass concentration from the aerosol extinction coefficient and the RH over the study period. The bias between the calculated PM$_{2.5}$ mass concentration and the SPS30-measured PM$_{2.5}$ mass concentration (bias = calculated value - measured value) versus the measured PM$_{2.5}$ is shown in Fig. 2(a). The MAE ($\text{MAE} = |\text{calculated value} - \text{measured value}|$) increased gradually as the PM$_{2.5}$ mass concentration increased. The errors show a clear correlation with PM$_{2.5}$. Increases in RH and pollutant emissions are often accompanied by an increase in the mass concentration of PM$_{2.5}$. RH and anthropogenic activities can affect aerosol composition and optical properties. A suitable empirical parameter was selected as a fixed value based on the dry aerosol particle size distribution. However, the aerosol optical parameters may change, and the calculated value of PM$_{2.5}$ may be too small owing to deviations in parameter selection. Moreover, changes in composition affect the hygroscopic properties of aerosols. Perhaps the calculated value of $f(RH)$ obtained using the aerosol size distribution for 5 consecutive days in autumn was greater than the average value of $f(RH)$ throughout the year, resulting in a decrease in the calculated value of PM$_{2.5}$.

The priori assumption of a constant of the extinction exponent “a” in Eqs. (9) and (12) was further refined or corrected because it may change with the aerosol type and humidity. According to the literature, RH and aerosol components together affect the aerosol extinction coefficient,
visibility, mass density, and refractive index (Qu et al., 2015). Additionally, aerosol types vary with
the emission sources and RH; therefore, a modification of the extinction exponent “a” is considered
to improve the PM$_{2.5}$ and extinction relationship. In achieving this objective, Eq. (12) can be
rewritten as

$$C_{PM2.5} = \frac{0.195}{3.302 \left(\frac{RH}{100}\right)^{7.211} + 1} \cdot (1000 \cdot \sigma_{RH})^{(0.96+\Delta a)}$$  \hspace{1cm} (13)

where $\Delta a = A(RH) + A(\sigma_{RH})$ represents total correction factors, $A(RH)$ represents the
correction factor related to RH, and $A(\sigma_{RH})$ represents the correction item relevant to aerosol
extinction and type uncertainty. These were estimated using the following steps:

1. The $\Delta a$ was calculated from Eq. (13) when the measured PM$_{2.5}$ mass concentration, RH,
   and aerosol extinction coefficient data were substituted into Eq. (13).

2. $\Delta a = A(RH)$ was obtained by neglecting $A(\sigma_{RH})$ when RH was high. This result is based
   on the reasonable assumption that $A(RH)$ generally plays a significant role when the RH is $>$60%.
   Fig. 3(a) shows the relationship between the mean $A(RH)$ and RH, averaged at
   2% RH intervals. The correlation is greatly fitted using a normal distribution, as shown in
   Eq. (14).

3. $A(\sigma_{RH})$ was calculated from $A(\sigma_{RH}) = \Delta a - A(RH)$ when RH was low. The
   relationship between the mean $A(\sigma_{RH})$ and the extinction coefficient was obtained, as
   shown in Fig. 3(b), averaged at 0.04 km$^{-1}$ extinction coefficient intervals. The fitted
   correlation (15) between $A(\sigma_{RH})$ and the extinction coefficient is the sum of the two
normal distributions.

\[ A(RH) = 0.1541 \exp \left( -\left( \frac{RH-120.9}{38.89} \right)^2 \right) \]  

(14)

\[ A(\sigma_{RH}) = 0.02435 \exp \left( -\left( \frac{\sigma_{RH}1000-206.1}{101.8} \right)^2 \right) - 0.1079 \exp \left( -\left( \frac{\sigma_{RH}1000-2558}{826.4} \right)^2 \right) \]  

(15)

Finally, the refined analytical expression or function for the dry PM$_{2.5}$ mass concentration, aerosol extinction coefficient, and hygroscopic factor was obtained as follows:

\[ C_{PM2.5} = \frac{0.195}{3.302 \left( \frac{RH}{100} \right)^{7.211}} \cdot (1000 \cdot \sigma_{RH})^{(0.96+\Delta a)}, \]  

(16)

where \( \Delta a = 0.1541 \exp \left( -\left( \frac{RH-120.9}{38.89} \right)^2 \right) + 0.02435 \exp \left( -\left( \frac{\sigma_{RH}1000-206.1}{101.8} \right)^2 \right) - 0.1079 \exp \left( -\left( \frac{\sigma_{RH}1000-2558}{826.4} \right)^2 \right) \). The extinction coefficient is in units of km$^{-1}$, and $C_{PM2.5}$ is in units of $\mu g \ m^{-3}$.  

Using Eq. (16), the PM$_{2.5}$ mass concentrations can be derived from the aerosol extinction coefficients and RH. Compared to the linear relationship expression used in the past (Tao et al., 2016, Ma et al., 2020), this new expression takes into account not only the variations in relative humidity affect optical properties, but also strives to quantify the impact of different pollution levels on optical properties by using the extinction coefficient. This represents an optimization in expressing the relationship between these variables. Next, we evaluated the derived PM$_{2.5}$ values by comparing them with the observed PM$_{2.5}$ in 2020 and January 2021, and the potential errors or uncertainties are discussed below.

4.3 Error Analysis
The PM$_{2.5}$ mass concentrations can be estimated or calculated using Eqs. (12) and (16), respectively, using observed aerosol extinction coefficients and RH data. The calculated values were compared with the measured values for PM$_{2.5}$, as shown in Fig. 2(b) and (c). The results indicate that the derived PM$_{2.5}$ from Eq. (16) was in better agreement with the measurements by showing lower RMSE and MAE, and a stronger correlation coefficient with the observed PM$_{2.5}$.

Fig. 4(a) displays the boxplots for the mean differences between the simulated PM$_{2.5}$ from Eqs. (12) and (16) and the measured PM$_{2.5}$ under various RH segments (1-year data). The mean errors demonstrate that Eq. (16) yields closer matches with the actual measurements than Eq. (12), particularly when the RH is >60%. The range of deviation increased as the RH increased. When the RH was >80%, the deviation range increased significantly. Considering the influence of Eq. (16), the error in humidity correction was more significant when the RH was >80%. However, from the perspective of environmental impact, water vapor directly affects visibility when RH is >80%, increasing the complexity of the atmospheric environment. The formula used to convert the visibility into aerosol extinction coefficients may produce apparent errors.

Fig. 4(b) shows a statistical plot of the variation in the deviation values of the calculated PM$_{2.5}$ by Eq. (16) with visibility. These results demonstrate that when visibility decreases, the bias range of PM$_{2.5}$ broadens. Large PM$_{2.5}$ concentrations often result in a reduction in visibility. In this case, the bias was increased owing to the increased PM$_{2.5}$ concentration. In the Supplementary material (Fig. S2), we make the error analysis of the upper limit of PM$_{2.5}$ mass concentration is best not to exceed 150 µg m$^{-3}$ because the error becomes large in heavy pollution weather. However, large RH also contributes to a reduction in visibility. These results were combined with those shown in Fig.
4(b). Under high-humidity conditions, the direct extinction effect of water vapor on the atmosphere leads to a significant error in converting visibility into aerosol extinction coefficients, which increases the error of the estimated PM$_{2.5}$.

Typically, the interaction between increasing pollution and humidity decreases the accuracy of the PM$_{2.5}$ estimate. The expression in Eq. (16) was refined using the RH and pollution levels; thus, the calculated results are consistent. Furthermore, the standard deviation of particle size distributions may result in some uncertainty on the expression of PM$_{2.5}$ and extinction coefficient (Eq. (12) and Eq. (16)). The constant “b” is calculated with the average particle size distribution (Fig. 1). If aerosol particle diameters are out of the range of OPC-N3, Eq. (12) or Eq. (16) will not work or the constant “b” needs to be changed.

Fig. 5 shows the temporal variation and comparison of the estimated PM$_{2.5}$ from Eq. (16), with observations during the four seasons of 2020. Overall, they showed good consistency in temporal variability and PM$_{2.5}$ magnitude. Strong linear correlations for all seasons are indicated with $R^2 > 0.8$. Eq. (16) achieved the best performance in autumn (RMSE=6.0132 $\mu$g m$^{-3}$), followed by spring (RMSE=7.8997 $\mu$g m$^{-3}$) and summer (RMSE=7.786 $\mu$g m$^{-3}$), which were similar, and winter (RMSE=13.4028 $\mu$g m$^{-3}$), which was the worst.

The absolute error between the SPS30-sensor measurements and station data was large at PM$_{2.5}$ mass concentrations greater than 100 $\mu$g m$^{-3}$. Due to high PM$_{2.5}$ mass concentrations and aerosol properties changing significantly in winter, the RMSE and MAE values were relatively high. High errors in the estimated PM$_{2.5}$ result from heavy pollution conditions. The lowest RMSE was achieved in autumn because the aerosol size distribution data were collected in autumn.
Autumn is characterized by clear, sunny days. Thus, other meteorological factors may produce fewer errors in PM$_{2.5}$ estimated from the aerosol extinction coefficient. The assumption of pertinent physical parameters may result in different degrees of error in different seasons owing to the seasonal characteristics of the aerosols.

4.4 Validation With TEOM-measured PM$_{2.5}$

To verify the applicability of the proposed method in Eq. (16), the derived PM$_{2.5}$ from Eq. (16) was compared with the independently measured PM$_{2.5}$ from the TEOM sampler (1274A station) near our study site. Comparisons between the two cases, in 2020 and January 2021, are shown in Figs. 6 and 7, respectively.

As shown in Fig. 6(a), we selected a case comparison of PM$_{2.5}$ on a sunny, clear day, September 6, 2020. The meteorological visibility and relative values are shown in Fig. 6. The comparisons of PM$_{2.5}$ indicate good consistency and similar temporal variability. RH was generally <70%. Visibility was approximately 20 km in the early morning but improved at noon and in the afternoon. A good agreement was achieved between the calculated and ground-measured PM$_{2.5}$ values with the daily MAE=3.93 µg m$^{-3}$ over the entire day. The MAE between the calculated values and TEOM observations (1274A station) was 6.61 µg m$^{-3}$.

Another moderate air pollution day (December 20, 2020) is shown in Fig. 6(b). The increase in the RH at night aggravates the polluted conditions. A good agreement is indicated between the calculated and ground-measured results with the daily MAE=9.21 µg m$^{-3}$ over the entire day. The MAE between the calculated values and the TEOM sampler data (1274A station) was 10.35 µg m$^{-3}$.
To verify the applicability of Eq. (16), we evaluated the derived PM$_{2.5}$, using the TEOM-measured data for January 2021. We first calculated PM$_{2.5}$ mass concentration using Eq. (16), aerosol extinction coefficient, and RH. Next, the calculated values of PM$_{2.5}$ mass concentration were compared with the measurements from the SPS30 sensor and the TEOM sampler at station 1274A. To avoid the influence of weather and other uncertainties, data from the periods of rain and snowfall were excluded. In addition, data with PM$_{10}$ less than the PM$_{2.5}$ mass were also removed. Hourly mean temperature, humidity, and wind speed data are shown in Fig. S4. The comparison results are shown in Fig. 7(a). The derived values (orange line) showed good agreement with the ground-measured values (black line from SPS30 and green line from the TEOM sensor at 1274A station data). In particular, they exhibited similar temporal variability and magnitude. Fig. 7(b) displays the correlation and linear regression between the derived PM$_{2.5}$ and the TEOM-measured PM$_{2.5}$ at station 1274A. A strong linear correlation was indicated, with a correlation coefficient of $R^2=0.76$ and MAE=8.8688 µg m$^{-3}$. The long-term validation results are shown in Supplemental Fig. S3, where we compared the 5-month (January, July, August, September, and October 2021) calculated values and the TEOM sampler measurement values. Good agreement was achieved with good correlation ($R^2=0.68$) and MAE=8.5856 µg m$^{-3}$. These comparisons further indicated that the analytical expression in Eq. (16) was applicable to different seasons in the study area.

### 4.5 Lidar Application Examples

The regional vertical distribution of PM$_{2.5}$ concentrations was derived by applying the CALIPSO-retrieved aerosol extinction profiles using Eq. (16). The 532 nm extinction coefficient and RH were from the CALIOP level-2 product (The vertical distribution of the 532 nm extinction...
coefficient is sourced from Version 4.2 5-km aerosol profile product. The vertical distribution of RH data is sourced from the MERRA-2 data.). Four case studies are presented in which CALIPSO passed over our study site: Hefei City and Anhui Province at 2:30 am Beijing time (BJT) on March 15, 2020; BJT at 2:50 am on November 12, 2020; BJT at 2:15 am on February 5, 2020; and BJT at 2:25 am on December 24, 2020. Fig. 8 (a) shows the CALIPSO ground track. To evaluate the derived PM$_{2.5}$ on the regional scale, we selected the ground TEOM-measured PM$_{2.5}$ at air quality stations in several major cities close to the CALIPSO tracks. The major cities are labeled Anqing (AQ), Luan (LA), Hefei (HF), Huainan (HN), Bengbu (BB), Fuyang (FY), Huangshan (HS), Xucheng (XC), Wuhu (WH), Maanshan (MAS), and Suzhou (SZ) in Figs. 8(a) and 9. Fig. 8 (b) shows the correlation between the estimated near-surface PM$_{2.5}$ from the CALIPSO extinction coefficients and the ground TEOM-measured values. A strong correlation was observed with a correlation coefficient of $R^2=0.94$ and a linear slope of 0.96, which implies a reliable estimate of PM$_{2.5}$. This initial comparison appears promising; however, the data points in the scatterplot are limited. An analysis using a greater number of data points than we used would have strengthened the statistical robustness.

Fig. 9 shows the vertical profiles of the CALIPSO aerosol extinction coefficients (Fig. 9a-d) and the corresponding derived PM$_{2.5}$ (Fig. 9e-h) in Anhui Province. The range-resolved spatial variation in PM$_{2.5}$ was observed on a regional scale. As shown in Fig. 9(c), the estimated PM$_{2.5}$ concentrations in the 30°–31°N latitude zone in southern Anhui (e.g., Anqing) were relatively higher than those in other regions. Fig. 9(f) shows high PM$_{2.5}$ concentrations below 1 km altitude from south to north Anhui on February 5, 2020 (winter). The near-surface PM$_{2.5}$ was estimated. The
values ranged from 80 to 160 µg m\(^{-3}\), exceeding China’s air quality standard (daily average 35 µg m\(^{-3}\) for Class 1 and 75 µg m\(^{-3}\) for Class 2, GB-3095 2012) and indicating a regional-scale air pollution event. On this day, the ground TEOM-measured PM\(_{2.5}\) values were 80 µg m\(^{-3}\) (AQ), 102 µg m\(^{-3}\) (LA), 94 µg m\(^{-3}\) (HF), 128 µg m\(^{-3}\) (HN), and 117 µg m\(^{-3}\) (BB). By contrast, the PM\(_{2.5}\) concentrations were relatively low on November 12 (fall), as shown in Fig. 9(g). Some near-surface polluted areas are indicated on December 24 (winter) in Fig. 9(h). The uncertainty analysis with CALIPSO/CALIOP reported uncertainties in the aerosol extinction coefficients (Figs. S5 and S6). When aerosol extinction coefficients are large at the near-surface, the uncertainties of aerosol extinction coefficients are relatively large, resulting in large uncertainty in the mass concentration of PM\(_{2.5}\). Under the same conditions, when the humidity range was > 60%, the error value of the PM\(_{2.5}\) mass concentration was significantly reduced (Fig. S7). Despite certain uncertainties, there was still a good prediction of air quality.

5 CONCLUSIONS

This study proposes a novel analytical expression for the PM\(_{2.5}\) mass concentration, aerosol extinction coefficient, and hygroscopic factor based on Mie theory and observation data analysis. The aerosol size distribution, PM\(_{2.5}\), visibility, and relative humidity (RH) in 2020 were used to refine the exponential factors relevant to the RH, aerosol type, and aerosol extinction coefficient uncertainties. By applying the analytical function, the derived PM\(_{2.5}\) from the aerosol extinction coefficient indicates good consistency with the independently observed PM\(_{2.5}\), showing similar temporal variation and a strong correlation coefficient in January 2021. The regional and vertical
distribution of PM$_{2.5}$ was reasonably derived from the NASA space-borne lidar CALIPSO/CALIOP aerosol extinction product in Anhui Province, China. The main conclusions are as follows:

(1) On the basis of Mie theory and the definition of PM$_{2.5}$ mass concentration, a non-linear exponential relationship among the PM$_{2.5}$ mass concentration, aerosol extinction coefficient, and RH was derived.

(2) The measurements of aerosol size distribution, PM$_{2.5}$, and RH were used to refine the extinction exponent and hygroscopic factor $f$(RH). The accuracy of the derived PM$_{2.5}$ is much improved compared with the observed PM$_{2.5}$ in the year 2020.

(3) The simulated PM$_{2.5}$ from aerosol extinction was further validated using tapered element oscillating microbalance (TEOM) measured PM$_{2.5}$ values in January 2021 by showing good consistency. Severe PM$_{2.5}$ pollution and high RH may result in large errors owing to the complex variability and uncertainty of aerosol type, size distribution, visibility, and aerosol extinction coefficient.

(4) The spatial and range-resolved distributions of PM$_{2.5}$, derived from the NASA/CALIPSO-observed aerosol extinction profiles, showed good agreement with ground TEOM-measured PM$_{2.5}$ values in a few major cities near the CALIPSO tracks.

This analytical expression was applicable within the error range. Although the heterogeneity of the spatial and temporal variability of aerosols may lead to a bias in applicability, the prediction accuracy can be improved by correcting the exponential factors with locally observed data.

The main limitation of this study is that the hygroscopic properties of different aerosol types or compositions can affect the accuracy of the estimated PM$_{2.5}$. Notably, long-term observations
may help differentiate aerosol types according to the season or weather. Thus, in further research, we plan to identify the aerosol types and improve the accuracy of the PM$_{2.5}$ estimate for the different types of PM$_{2.5}$. The findings of this study can be used in lidar remote sensing techniques to monitor the spatial distribution of PM$_{2.5}$ mass concentrations and help understand the transport and dispersion processes of aerosols.

**ACKNOWLEDGMENTS**

This work was supported by the HFIPS Director's Fund (YZJJ202303-TS), the Open Project of Advanced Laser Technology Laboratory of Anhui Province (AHL2021ZR01), and the National Natural Science Foundation of China (U20A20214). The authors would like to thank the data product from National Environmental Monitoring Center and NASA. The authors also gratefully acknowledge the Anhui Meteorological Observation Technical Center.

**DISCLAIMER**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**REFERENCES**


Fig. 1. (a) Hourly average aerosol volume size distribution with different relative humidity (RH) from October 9 to 13, 2020, at Hefei. (b) Relationship between $f(RH)$ and RH. Star points are $f(RH)$ values calculated from the average size distribution at different relative humidity levels with 2% humidity intervals. The solid red line indicates the fitted curve.
Fig. 2. (a) Error or bias between the estimated PM$_{2.5}$ and the measured PM$_{2.5}$ mass concentrations (blue dots) and the mean absolute error (MAE, triangles, calculated at intervals of 20 µg m$^{-3}$). (b) and (c): Comparison of the correlation between calculated and measured PM$_{2.5}$ mass concentration values. (b) Computed result of Eq. (12), (c) Computed result of Eq. (16). The dashed black line represents the 1:1 line, and the red line represents the best-fit line from linear regression. R represents the determination coefficient, RMSE represents the root mean square error, MAE represents the mean absolute error, and N represents the number of samples. The equation with the terms Y and X represents the fitting relationship between the measured and calculated PM$_{2.5}$ mass concentration values. Points are color-coded by the number of data points per 1 µg m$^{-3}$ bin. (Fig.2(c), the slope is closer to 1, and the RMSE and MAE decrease).
Fig. 3. (a) Relationship between $A(RH)$ and RH and (b) relationship between $A(\sigma_{RH})$ and $\sigma_{RH}$. Black star points are calculated mean values. Relative humidity interval, 2%; extinction coefficient interval, 0.04 km$^{-1}$. 
Fig. 4. (a) Comparison of PM$_{2.5}$ bias values with relative humidity before and after correction. Boxplots of resulting bias (y-axis) for different RH (x-axis) ranges with an interval of 10%. The boundary of each box at the bottom designates the 25$^{th}$ percentile. The closest boundary of the box to the top designates the 75$^{th}$ percentile. The median is shown with a line within each box. Error bars above and below the boxes indicate the 90$^{th}$ and 10$^{th}$ percentiles of the parameters. (b) The plot of PM$_{2.5}$ deviation values with visibility (VIS). Boxplots of resulting bias (y-axis) for different VIS (x-axis) ranges with a VIS interval of 5 km.
Fig. 5. Comparison between the derived PM$_{2.5}$ and the SPS30 measured values for the four seasons of 2020. (a) Spring, (b) Summer, (c) Autumn, and (d) Winter.
Fig. 6. Comparative analysis of calculated and measured PM$_{2.5}$ mass concentration values for two typical weather conditions. (a) September 6, 2020, was a clear day; (b) there was light-moderate pollution on December 20, 2020.
Fig. 7. (a) Comparison of the derived PM$_{2.5}$ mass concentration values with the ground-measured values. The black, orange, and green dotted lines represent the ground-measured value from SPS30 and the calculated and TEOM-measured value at national control station 1274A, respectively, (b) Correlation and regression analysis between the derived PM$_{2.5}$ and the TEOM-measured values at national control station 1274A.
Fig. 8. (a) CALIOP roadmap through Anhui Province; (b) Correlation between the CALIOP-derived near-surface PM$_{2.5}$ and the TEOM-measured PM$_{2.5}$ at the air quality stations.
Fig. 9. (a)-(d) CALIOP-retrieved aerosol extinction coefficient at 532 nm on March 15, 2020; February 5, 2020; November 12, 2020; and December 24, 2020, respectively. (The horizontal dash-line curve is the surface elevation from the CALIPSO product and the vertical dash lines indicate the positions for a few major cities near the CALIOP track). (e)-(h) CALIOP-derived PM$_{2.5}$ concentrations corresponding to Fig. 9 (a)-(d), respectively.