Supplementary material:

Analysis of non-linear relationship of PM$_{2.5}$ mass concentration with aerosol extinction coefficient and RH in Hefei, China

Yinan Chen$^{1,2,3}$, Shiguo Zhang$^{1,2,3,4}$, Yonghua Wu$^{5}$, Kee Yuan$^{1,3}$, Jian Huang$^{1,3}$, Dongfeng Shi$^{1,3}$, Shunxing Hu$^{1,3*}$

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

Corresponding author: Shunxing Hu (sxhu@aiofm.ac.cn)
S1 Methodology

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

$$\sigma_{RH} = \int_0^\infty \pi Q_{ext}(m, r, \lambda) r^2 n(r)_{RH} dr \approx \pi Q_{ext} \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr$$  \hspace{1cm} (S1)

Where $r$ denotes the particle radius (unit: $\mu$m). $m$ represents the complex refractive index, $Q_{ext}$ represents the particle extinction efficiency factor for a single particle at $m$, $r$, and $\lambda$. $n(r)_{RH} = dN/d1gr$ denotes the particle size distribution under the relative humidity (RH) (unit: $cm^{-3}$). $Q_{ext}$ represents the overall average extinction efficiency factor (Koelemeijer et al., 2006, Hoff et al., 2009).

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

$$C_{PM2.5} = \int_0^{1.25} \frac{4\pi \rho}{3} r^3 n(r)_{dry} dr \approx 4\pi \bar{\rho} \int_{r_{min}}^{r_{max}} r^3 n(r)_{dry} dr$$  \hspace{1cm} (S2)

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 mean effective density (Koelemeijer et al., 2006, Hoff et al., 2009).

Taking the logarithm of both sides of Eq. (S1), we get:

$$ln \sigma_{RH} = ln[\pi Q_{ext}] + ln \left[ \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr \right]$$  \hspace{1cm} (S3)

Then, taking the logarithm of both sides of Eq. (S2), we get:

$$ln C_{PM2.5} = ln \left( \frac{4\pi \bar{\rho}}{3} \right) + ln \int_{r_{min}}^{1.25} r^3 n(r)_{dry} dr$$  \hspace{1cm} (S4)

Eq. (3) multiplying by the coefficient $a$, $a = log(\pi Q_{ext}) \left( \frac{4\pi \bar{\rho}}{3} \right)$ is a constant, Eq. (S3) becomes:

$$a \ ln \sigma_{RH} = a \ ln[\pi Q_{ext}] + a \ ln \left[ \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr \right]$$  \hspace{1cm} (S5)

We subtract Eq. (S5) from Eq. (S4) (the left and right sides of the two equations are subtracted, respectively), and the result is as follows:

$$ln C_{PM2.5} - a(ln \sigma_{RH}) = ln \left( \frac{4\pi \bar{\rho}}{3} \right) + ln \int_{r_{min}}^{1.25} r^3 n(r)_{dry} dr - a \ln \left[ \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr \right]$$  \hspace{1cm} (S6)

The value of “$a$” is $log(\pi Q_{ext}) \left( \frac{4\pi \bar{\rho}}{3} \right)$, thus $ln \left( \frac{4\pi \bar{\rho}}{3} \right) - a(ln[\pi Q_{ext}]) = 0$, Then, Eq. (S6) becomes:

$$ln C_{PM2.5} - a(ln \sigma_{RH}) = ln \int_{r_{min}}^{1.25} r^3 n(r)_{dry} dr - a \ln \left[ \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr \right]$$  \hspace{1cm} (S7)

According to the logarithmic algorithm, Eq. (S6) becomes:

$$ln C_{PM2.5} - ln(\sigma_{RH})^a = ln \int_{r_{min}}^{1.25} r^3 n(r)_{dry} dr - ln \left[ \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr \right]^a$$

$$ln \left( \frac{C_{PM2.5}}{(\sigma_{RH})^a} \right) = ln \left( \frac{\int_{r_{min}}^{1.25} r^3 n(r)_{dry} dr}{\left[ \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr \right]^a} \right)$$  \hspace{1cm} (S8)

And making $b = \frac{\int_{r_{min}}^{1.25} r^3 n(r)_{dry} dr}{\left[ \int_{r_{min}}^{r_{max}} r^2 n(r)_{RH} dr \right]^a}$ and removing the logarithm from Eq. (S8), then we get
the analytical relationship between the PM$_{2.5}$ mass concentration and aerosol extinction ($\sigma_{RH}$) in the following:

$$C_{PM_{2.5}} = b \sigma_{RH}^a$$  \quad (S9)

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 PM$_{2.5}$ mass concentration must be considered.

$$b = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{dry}} dr}{\left(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{RH}} dr\right)^a} = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{dry}} dr}{\left(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{RH}} dr\right)^a} = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{dry}} dr}{f(RH) \left(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{RH}} dr\right)^a}$$  \quad (S10)

Where $b_{\text{dry}} = \frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{dry}} dr}{\left(\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{RH}} dr\right)^a}$, the unit of $b$ is $\mu g \cdot \left(\frac{1}{m}\right)^{-a}$; $f(RH) = \left(\frac{\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{RH}} dr}{\int_{r_{\text{min}}}^{r_{\text{max}}} r^2 n(r)_{\text{RH}} dr}\right)^a$.

$b_{\text{dry}}$ is a constant associated with dry aerosols and $f(RH)$ 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:

$$C_{PM_{2.5}} = \frac{b_{\text{dry}} \sigma_{RH}^a}{f(RH)}$$  \quad (S11)

Among them: $a$ is a constant, the unit of $b$ is $\mu g \cdot \left(\frac{1}{m}\right)^{-a}$, the unit of $\sigma_{RH}$ is $m^{-1}$, the unit of $C_{PM_{2.5}}$ is $\mu g m^{-3}$.

**S2 Results and Discussion**

<table>
<thead>
<tr>
<th>RH (%)</th>
<th>$V_{PM_{2.5}}/V_{0PM_{2.5}}$ (RH &lt;60%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RH&lt;60%</td>
<td>1.00</td>
</tr>
<tr>
<td>60%&lt;RH&lt;=70%</td>
<td>1.42</td>
</tr>
<tr>
<td>70%&lt;RH&lt;=80%</td>
<td>1.68</td>
</tr>
<tr>
<td>80%&lt;RH&lt;=90%</td>
<td>2.95</td>
</tr>
<tr>
<td>RH&gt;90%</td>
<td>5.90</td>
</tr>
</tbody>
</table>

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. When the PM$_{2.5}$ mass concentration was >100 $\mu g m^{-3}$, a significant overestimation was observed. This result was consistent with the precision of the instrument, which might be attributed to the limited algorithm (the accuracy of measurements from 100 to 1000 $\mu g m^{-3}$ was ±10%).
Fig. S1 Comparison of SPS30-measured PM$_{2.5}$ mass concentration values with the data from the nearest national control station (1274A).

Fig. S2 shows a statistical plot of the variation in the deviation values of the calculated PM$_{2.5}$ by Eq. (16) with measured PM$_{2.5}$. According to Fig. S2, results are great under excellent and good weather conditions (PM$_{2.5} < 75$ µg m$^{-3}$). In light and moderate pollution weather conditions (75 µg m$^{-3} <$ PM$_{2.5} < 150$ µg m$^{-3}$), the results are good in most cases. However, in heavy pollution weather (PM$_{2.5} > 150$ µg m$^{-3}$), there is an obvious underestimation, and the MAE > 20 µg m$^{-3}$. In summary, 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.

Fig. S2. The error of PM$_{2.5}$ deviation values versus the measured PM$_{2.5}$ mass concentration. Boxplots of resulting bias (y-axis) for different measured PM$_{2.5}$ mass concentration (x-axis) range with a PM$_{2.5}$ mass concentration interval of 10 µg m$^{-3}$. The boundary of each box at the bottom designates the 25th percentile. The closest boundary of the box to the top designates the 75th percentile. The median is shown with a line within each box. Error bars above and below the boxes indicate the 90th and 10th percentiles of the parameters.

To further verify the applicability of Eq. (16), we evaluated the derived PM$_{2.5}$ values with the TEOM-measured data in January, July, August, September, and October 2021. The data that reach the measurement limits of the instrument need to be excluded. To avoid the influence of weather
and other uncertainties, the data in the periods of rain and snowfall are excluded. In addition, the data with the PM$_{10}$ less than the PM$_{2.5}$ mass are also excluded.

We first calculated PM$_{2.5}$ mass concentration using Eq. (16) and the aerosol extinction coefficient and relative humidity. Then, the comparison results between the calculated values and TEOM sampler measuring values are shown in Fig. S3. 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.

According to the comparison results, there may be significant errors in some extreme weather conditions, such as windy, dusty, and weather with obvious external pollution sources. These uncertain factors may change the size distribution and optical properties of the aerosol. The factors that affect the error will be the content of our in-depth research.

![Fig. S3. Correlation and regression analysis between the derived-PM$_{2.5}$ and the TEOM-measured values at the national control station 1274A (data in January, July, August, September, and October 2021).](image-url)
Fig. S4. Hourly mean data of (a) temperature, (b) humidity, and (c) wind speed over time in January 2021 (Fig. 7(a)). The temperature and humidity data were collected by the HMP155A Humidity and Temperature Sensors, while the meteorological station (Hefei Meteorological Science and Technology Park Station) provided the wind speed data.

We made the uncertainty analysis with the CALIPSO/CALIOP reported uncertainties of aerosol extinction coefficients, and the results for the uncertainties of CALIOP-derived PM$_{2.5}$ concentrations are shown in Fig. S5. When aerosol extinction coefficients are large at near-surface, the uncertainties of aerosol extinction coefficients are relatively large, resulting in a large uncertainty in the mass concentration of PM$_{2.5}$. The results of a relative error (the ratios of PM$_{2.5}$ uncertainties to the estimated PM$_{2.5}$ concentrations) are shown in Fig. S6. Despite certain uncertainties, there was still good prediction for air quality.

![Uncertainty Analysis](image)

Fig. S5. (a)-(d) CALIOP-retrieved uncertainty of aerosol extinction coefficient at 532 nm on March 15, 2020; February 5, 2020; November 12, 2020; and December 24, 2020, respectively. (The horizontal dash-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) Uncertainty of derived-PM$_{2.5}$ concentrations corresponding to Fig. S5 (a)-(d), respectively.
Fig. S6. (a)-(d) Relative error (the ratios of PM$_{2.5}$ uncertainties to the estimated PM$_{2.5}$ concentrations) on March 15, 2020, February 5, 2020, November 12, 2020, and December 24, 2020, respectively.

The error transfer value caused by the extinction coefficient is calculated from Equation 16. Under the conditions of different relative humidity, the variation of the error value of PM$_{2.5}$ mass concentration ($\Delta PM_{2.5} = C'_{PM_{2.5}} - C_{PM_{2.5}}$) with the error value of extinction coefficient ($\Delta \sigma_{RH}$) is shown in Fig. S7.

Where:

$$C_{PM_{2.5}} = \frac{0.195}{3.302 \left( \frac{RH}{100} \right)^{0.2711} + 1} \cdot (1000 \cdot \sigma_{RH})^{0.96 + \Delta a} \tag{S12}$$

where: \( \Delta a = 0.1541 \exp \left(-\frac{RH-120.0}{10.00}\right) + 0.02435 \exp \left(-\frac{\left(\frac{RH-100-206.3}{101.8}\right)^2}{826.4}\right) - 0.1079 \exp \left(-\frac{\left(\frac{RH-100-255.5}{826.4}\right)^2}{826.4}\right) \).

$$C'_{PM_{2.5}} = \frac{0.195}{3.302 \left( \frac{RH}{100} \right)^{0.2711} + 1} \cdot (1000 \cdot (\sigma_{RH} + \Delta \sigma_{RH}))^{0.96 + \Delta a} \tag{S13}$$

where: \( \Delta a' = 0.1541 \exp \left(-\frac{RH-120.0}{10.00}\right) + 0.02435 \exp \left(-\frac{\left(\frac{RH+\Delta \sigma_{RH}-100-206.3}{101.8}\right)^2}{826.4}\right) - 0.1079 \exp \left(-\frac{\left(\frac{RH+\Delta \sigma_{RH}-100-255.5}{826.4}\right)^2}{826.4}\right) \).

The error value of PM$_{2.5}$ mass concentration increases with worse air quality. ($\Delta \sigma_{RH} < \sigma_{RH}$). Under the same conditions, when the humidity range is greater than 60%, the error value of the PM$_{2.5}$ mass concentration is significantly reduced.
Fig. S7. Under the condition of different relative humidity, the error value of PM$_{2.5}$ mass concentration varies with the error value of the extinction coefficient. (a) Clean: $\sigma_{RH} = 0.1 \text{ km}^{-1}$; (b) Average: $\sigma_{RH} = 0.5 \text{ km}^{-1}$; (c) Pollution: $\sigma_{RH} = 1 \text{ km}^{-1}$. 