



Contents lists available at ScienceDirect

## Remote Sensing of Environment

journal homepage: [www.elsevier.com/locate/rse](http://www.elsevier.com/locate/rse)

## Lidar-based daytime boundary layer height variation and impact on the regional satellite-based PM<sub>2.5</sub> estimate

Sijie Chen<sup>a</sup>, Bowen Tong<sup>a</sup>, Lynn M. Russell<sup>b</sup>, Jing Wei<sup>c</sup>, Jianping Guo<sup>d</sup>, Feiyue Mao<sup>e</sup>, Dong Liu<sup>f</sup>, Zhongwei Huang<sup>g</sup>, Yun Xie<sup>h</sup>, Bing Qi<sup>i</sup>, Huan Zhang<sup>i</sup>, Yingshan Sun<sup>a</sup>, Boran Zhang<sup>a</sup>, Chengkang Xu<sup>a</sup>, Lingyun Wu<sup>a</sup>, Dong Liu<sup>a,j,k,\*</sup>

<sup>a</sup> State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 310027, China

<sup>b</sup> Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

<sup>c</sup> Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA

<sup>d</sup> State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China

<sup>e</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

<sup>f</sup> Key Laboratory of Atmospheric Composition and Optical Radiation, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei, Anhui 230031, China

<sup>g</sup> Collaborative Innovation Center for West Ecological Safety (CIWES), College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China

<sup>h</sup> Fuyang Meteorological Bureau, Hangzhou 311499, China

<sup>i</sup> Hangzhou Meteorological Bureau, Hangzhou 310051, China

<sup>j</sup> Jiaxing Key Laboratory of Photonic Sensing & Intelligent Imaging, Jiaxing 314000, China

<sup>k</sup> Intelligent Optics & Photonics Research Center, Jiaxing Research Institute Zhejiang University, Jiaxing 314000, China

## ARTICLE INFO

Edited by Menghua Wang

## Keywords:

PM<sub>2.5</sub>  
MAIAC AOD  
Haze layer height  
Deep learning  
MPL

## ABSTRACT

Satellite-derived aerosol optical depth (AOD) is popularly used to infer ground-level PM<sub>2.5</sub> concentration due to its wide coverage. The fact that aerosols are largely confined in the atmospheric boundary layer makes boundary layer height (BLH) an important scale factor for AOD-based PM<sub>2.5</sub> estimates. Our recent ground-based lidar observations, nevertheless, indicate that aerosol particles are heterogeneously mixed within the boundary layer, and even frequently reside above BLH, forming the residual layers (RL)-like pattern. To better sort out the underlying mechanism behind the above-mentioned phenomenon and the impact on hourly ground-level PM<sub>2.5</sub> estimates from satellite-based AOD, we firstly propose a novel notion of haze layer height (HLH), which is calculated from Micro-Pulse Lidar (MPL) profile. Combined analysis of 3.5-year ground-based lidar profiles, CE-318 AOD, and PM<sub>2.5</sub> measurements show that the coefficient of determination ( $R^2$ ) between PM<sub>2.5</sub> and AOD normalized with HLH increases from 0.49 to 0.61 for 90% of the dataset. Second, we applied HLH to an Auto-encoder-based Deep Residual Network (ADRN) and tested the effect on satellite AOD-based PM<sub>2.5</sub> estimation within a 300 km range surrounding the MPL station. With the aid of the AOD imputation technique, a similar improvement of using HLH instead is found on the regional scale PM<sub>2.5</sub> estimation, which can be demonstrated by the comparison with air quality measurements and other machine learning models. The results show that using ADRN with HLH achieves the highest performance (mean  $R^2 = 0.87$ , RMSE = 10.12  $\mu\text{g}/\text{m}^3$ ) among 4 machine learning models. This new approach, largely combining active and passive remote sensing data through artificial neural networks (CAPTA), shows improved accuracy and coverage of hourly PM<sub>2.5</sub> estimation with aerosol vertical information and AOD calculation under clouds. In addition, further analysis showed that the average difference between morning and daily PM<sub>2.5</sub> concentration could equate to an accuracy of 0.19–2.57 yrs. in terms of life expectancy, indicating that our new approach to the determination of PM<sub>2.5</sub> from space sheds new insight into the assessment of the aerosol impact on public health.

\* Corresponding author at: State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 310027, China.

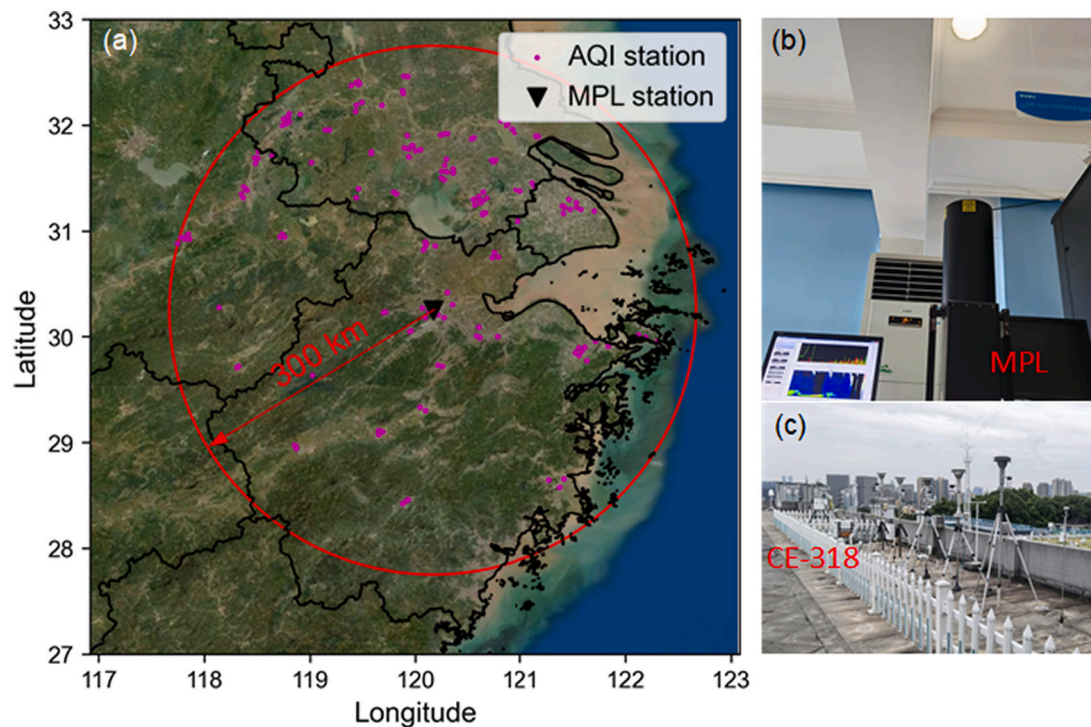
E-mail address: [liudongopt@zju.edu.cn](mailto:liudongopt@zju.edu.cn) (D. Liu).

<https://doi.org/10.1016/j.rse.2022.113224>

Received 28 February 2022; Received in revised form 31 July 2022; Accepted 14 August 2022

Available online 23 August 2022

0034-4257/© 2022 Elsevier Inc. All rights reserved.



**Fig. 1.** Location of 192 AQI stations (a) and key instruments at MPL station used in this work (b-c). The red circle in panel (a) marks the boundary of the study area. The MPL in panel (b) and CE-318 in panel (c) are located in the same building. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## 1. Introduction

Aerosols or airborne particulate matters (PMs) near the ground surface, especially for fine particles with an aerodynamic diameter  $<2.5 \mu\text{m}$  ( $\text{PM}_{2.5}$ ), are known to significantly affect the weather system and climate change on the Earth and bring adverse effects on public health (Bellouin et al., 2020; Liang et al., 2020). Led by the World Air Quality Index project, the air quality index (AQI) is measured from  $>30,000$  stations in 2000 major cities around the globe (<https://aqicn.org>), but the coverage is sparse in non-city areas. A variety of techniques have been developed to provide practical estimates of  $\text{PM}_{2.5}$  that supplement the insufficient in situ observations. Aerosol optical depth (AOD), a metric used to indicate aerosol loading in the vertical column, has been found to be able to be empirically converted into mass concentration at ground level. In recent decades, advanced space-borne sensors provide AOD measurements with broad spatial coverage as well as high spatial resolution, making it a widely used source in the AOD- $\text{PM}_{2.5}$  conversion (Hu et al., 2014; Lin et al., 2015; Wu et al., 2016; Zhang et al., 2018).

However, the inability of passive sensors to measure vertical profiles leads to a major difficulty in  $\text{PM}_{2.5}$  estimation. Since the measured AOD is an optical column property lacking information about vertical distribution, simply converting column AOD to surface  $\text{PM}_{2.5}$  might lead to significant errors (Lewis et al., 2010). A scaling parameter has to be assumed, typically referred to as the mixing layer height (MLH), which ideally equals the boundary layer height (BLH) (Emeis et al., 2008). The boundary layer is the lowest portion of the atmosphere where the exchange of mass, heat, and momentum mainly occur, making aerosol particles well-mixed inside the layer, below the free atmosphere (Li et al., 2018). In other words, BLH indicates the volume for the dispersion of aerosols concentrated near the surface in the vertical direction (Pal et al., 2015). BLH is a crucial factor for studies estimating surface  $\text{PM}_{2.5}$  concentration through either empirical or semi-empirical statistical methods (Tian and Chen, 2010; Sorek-Hamer et al., 2013; Song et al., 2014), chemical transport models (Xing et al., 2015; Geng et al., 2017; van Donkelaar et al., 2019), or machine learning approaches (Huang

et al., 2018; Wei et al., 2019; Li et al., 2020b; Li, 2020a, 2020b).

Lidar is an active remote sensing technique that measures the intensity of backscattered light as a function of range; ground-based MPL has been widely used for profiling the vertical structure of clouds and aerosols, forming networks that provide long-term observations at multiple sites around the globe (Berkoff, Welton et al., 2004; Huang et al., 2012; Cordoba-Jabonero et al., 2018). In addition, new lidar techniques have continued to be developed with more advanced systems or algorithms (Shen et al., 2020; Xiao et al., 2020; Wang et al., 2022). Lidar-based researches on the atmospheric vertical structure indicate that BLH is not always ideal for the scaling parameter (Chu et al., 2015; Reid et al., 2017; Zhao et al., 2018). For instance, when the daytime convective boundary layer (CBL) transitions to a shallow nocturnal boundary layer (NBL) at sunset, a significant amount of aerosols are frequently observed above BLH as a residual layer (RL) (Geiss et al., 2017). In other words, aerosols are not always confined in the boundary layer, nor are they always well-mixed throughout this layer.

To address this issue, a few prior studies have proposed a variety of correction methods. Chu et al. (2013) and their follow-up work (Chu et al., 2015) took the aerosols above BLH into account by calculating haze layer height (HLH), which is defined as the sum of BLH and an additional scale height, and achieved improvements in estimated  $\text{PM}_{2.5}$  concentration. Yet, due to the scarcity of coincident lidar observations and satellite measurements, the potential of improving AOD- $\text{PM}_{2.5}$  estimation using lidar profiles is rarely tested on a long-term basis. The inability of space-borne imagers such as MODerate resolution Imaging Spectroradiometer (MODIS) to resolve AOD when clouds are present adds to the difficulty of comparison (Levy et al., 2013). Despite their wide swaths, the effective AOD measurements are severely limited by cloud cover. In addition, related studies suggest relative humidity (RH), wind speed, and surface temperature must be considered in the AOD- $\text{PM}_{2.5}$  conversion (Song et al., 2014; Li et al., 2017).

New approaches using artificial intelligence (machine and deep learning) provides possible ways to solve complicated problems regarding aerosols, including both the missing AOD observations and

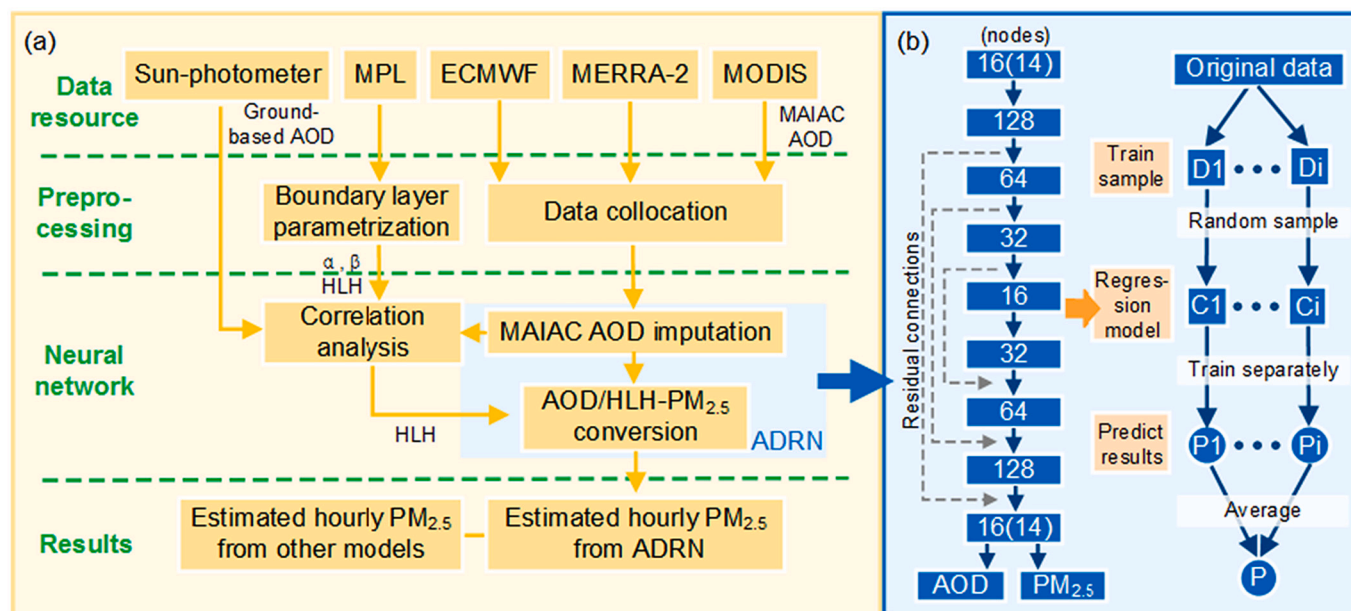


Fig. 2. (a) schematic diagram of the study and (b) framework of the autoencoder-based deep residual neural network.

the complicated non-linear relationship in the AOD- $PM_{2.5}$  conversion (Sun et al., 2019; Wang et al., 2019; Chen et al., 2020; Yan et al., 2020; Liu et al., 2021). Zang et al. (2019) used a generalized regression neural network model for estimating PM mass; in addition, it proved that the quality of AOD and the robustness of the interpolation algorithm affect the model accuracy. Wei et al. (2019) developed a space-time random forest (STRF) model to estimate 1 km resolution  $PM_{2.5}$  with considerable accuracy ( $R^2 = 0.85$ ,  $RMSE = 15.57 \mu g/m^3$ ) in China from the multiple implementation of atmospheric correction (MAIAC) AOD filled by a linear regression data integration approach; and then continuously improved the  $PM_{2.5}$  estimates using newly developed space-time extratrees (Wei et al., 2021a, 2021b) and space-time LightGBM (Wei et al., 2021a, 2021b) models. Li trained an autoencoder-based deep residual network (ADRN) with MAIAC AOD and other input variables; the resulting weekly AOD over California showed a high  $R^2 = 0.94$  in the independent test (Li et al., 2020a). The author's other work applied the same model to Beijing, China, and further estimated  $PM_{2.5}$ , which also achieved high  $R^2$  (0.82–0.97) compared to actual measurements (Li, 2020a, 2020b). Overall, deep learning models have successfully estimated  $PM_{2.5}$  on a daily, weekly, or larger time scale. Still, few of them have included the variables regarding the observation of boundary layer variation and estimate  $PM_{2.5}$  concentration on an hourly basis. These studies, focusing on daily average or satellite overpass (typically around noon), may lack information pertaining to morning or afternoon peak hours that  $PM_{2.5}$  concentration can be several times higher than the relatively well-mixed period around noon (Hasheminassab et al., 2014). Combining with rush hours, high  $PM_{2.5}$  concentration in those periods can cause significant exposure, leading to negative impact on human health, and thus must be further studied (Gupta and Elumalai, 2019).

This study proposes a method combining active and passive instruments through artificial neural networks (CAPTA). The traditional AOD/BLH- $PM_{2.5}$  estimation is improved by introducing a new formula for parametrizing HLH from coincident ground-based lidar observations. HLH derived from the lidar profile accounts for mixing conditions inside the layer and residual components above the layer. The local pattern of boundary layer variation in the Yangtze River Delta (YRD) region of China is analyzed in detail from 2017 to 2020. The proposed normalization of AOD by HLH is first tested through correlation analysis using ground-based coincident measurements at the MPL station. Further analysis on the regional scale is performed using MAIAC AOD

with imputation from the ADRN model. The accuracy of AOD imputation is examined through comparison with measurements from the CE-318 sun-photometer. The regional impact of AOD/HLH normalization is assessed for each AQI station and its distance to the lidar. Afterward, HLH is added to ADRN to produce hourly  $PM_{2.5}$  concentrations. The performance of trained ADRN is self-tested and compared with those derived from other machine learning models.

## 2. Data and methods

### 2.1. Study region and data sources

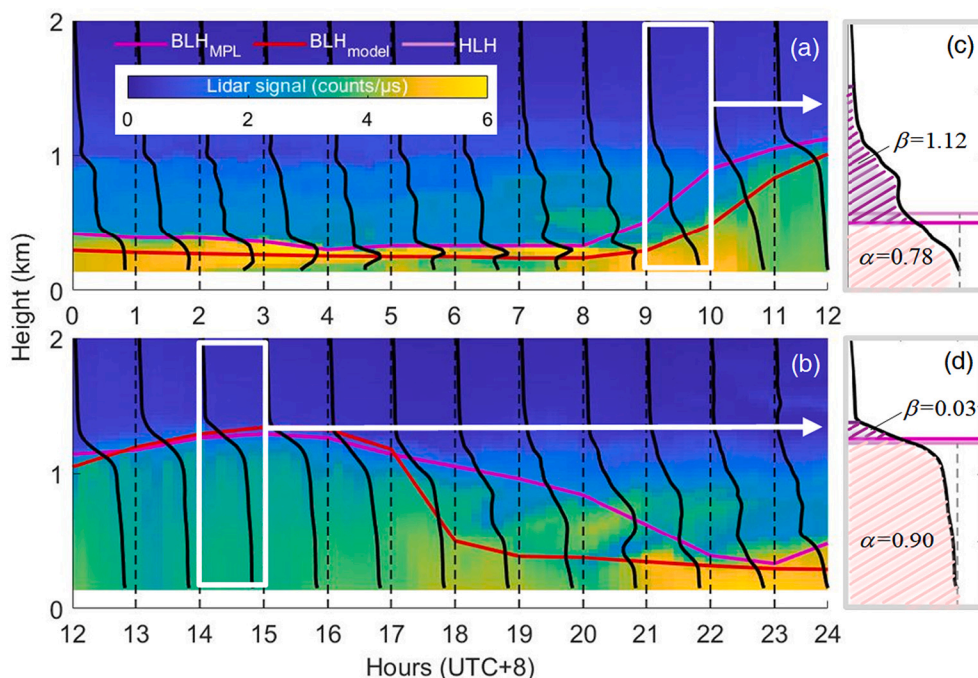
This study focuses on a 300 km radius region covering most of the population centers in the YRD (Fig. 1). The center of the studied region is Hangzhou (120.10°E, 30.14°N), the largest city of Zhejiang province. This region is far from the natural dust source in northwestern China and generally has calm weather due to the topography of a vegetated plain with hilly areas. Aloft aerosol layers from long-range transport are rarely observed and therefore not considered in this study, though trans-boundary transport might occur (Huang et al., 2020).

From January 2017 to May 2020, AOD from both Terra and Aqua MODIS sensors and a CE-318 sun-photometer (Fig. 1c),  $PM_{2.5}$  concentrations from 192 AQI stations, managed by the Ministry of Ecology and Environment of China (Fig. 1a), and atmospheric profiles from an MPL (Fig. 1b), as well as local meteorological data (such as RH and temperature), were collected at hourly resolution. Local time (UTC +8) is used in the following context. Data coverage (Table S1) and pre-processing are detailed in the supplement. Fig. 2 shows the schematic diagram of the study.

Since the MPL profile has a blind area (non-detectable range) from surface to 150 m, and model data indicate that BLH might be lower than that limit in early morning and evening hours, we focus on the measurements made between 8:00 and 19:00.

### 2.2. Boundary layer parameterization

Theoretically, BLH is frequently considered to be equal to MLH if all aerosols are confined within this homogeneously mixed layer, but such assumptions could be easily violated when the boundary layer is not uniformly mixed, or when a residual layer is present (Schneider and



**Fig. 3.** Times-series of the BLH observed on November 1st, 2019. MPL measured range-corrected signal (a-b) at 532 nm represents changes in aerosol loading. BLHs determined by different techniques are presented (MPL in pink, NWP in red, calculated HLH in light pink). Hourly averaged profiles are marked in black lines. Profiles at 9:00 (c) and 14:00 (d) local time are shown in detail, with correction coefficients  $\alpha$  and  $\beta$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Eixmann, 2002; Boyouk et al., 2010; Chu et al., 2015) Therefore, we derive HLH from BLH indicated by lidar profile and corrected for deviation from the ideal case.

The BLH from the MPL profile is calculated using the gradient method, which has been demonstrated to work well on range-corrected lidar signals (Tsaknakis et al., 2011; Nakouidi et al., 2019). The principle of the gradient method is to find the steepest gradient from the back-scattered light intensity profile, which happens at the transition between the boundary layer and the clean atmosphere above, where aerosol concentration sharply decreases. Technically, the BLH determined by the gradient method is at the center of the entrainment zone, where interference with the free troposphere has started. Note that inconsistencies in the definition of the BLH exist among different models and algorithms, which could lead to disagreement in the results. Before the method is applied, the profiles are screened to remove cloud contamination indicated by strong scattering signals ( $>10$  counts/ $\mu\text{s}$ ) and gradient much larger than the typical conditions between the boundary layer and the free troposphere.

Next, the correction coefficients are derived from the lidar profile. First, the degree of vertical mixing within the BLH ( $\alpha$ ) is defined as the ratio between the integral of lidar signal up to the retrieved BLH from MPL ( $\text{MPL}_{\text{BLH}}$ ) to that of the assumed well-mix condition:

$$\alpha = \frac{\int_0^{\text{BLH}_{\text{MPL}}} P(z) dz}{\int_0^{\text{BLH}_{\text{MPL}}} P_0 dz} \quad (1)$$

where  $P$  is the range-corrected signal, which has also been calibrated for overlap function,  $P_0$  is the signal at 150 m, the first value above the lidar blind area. This factor has not been considered in previous work by Chu et al. (2015), yet we think it is important to evaluate the mixing condition inside the boundary layer. Here we assumed that aerosols are well-mixed in the blind area (where we do not have lidar signal) but can be non-uniform between 150 m and BLH. Second, the amount of residuals above  $\text{MPL}_{\text{BLH}}$  is compressed within the boundary layer; the factor  $\beta$  is defined as the ratio between the integral of lidar signal in the residual layer to that in the boundary layer:

$$\beta = \frac{\int_{\text{BLH}_{\text{MPL}}}^{\text{clear}} P(z) dz}{\int_0^{\text{BLH}_{\text{MPL}}} P(z) dz} \quad (2)$$

where *clear* is the calculated baseline of the clean atmosphere above, defined by approaching (10% higher than) the minimum value of profile in the troposphere.

Thus, HLH in this work can be calculated from the BLH and correction coefficients using the following equation:

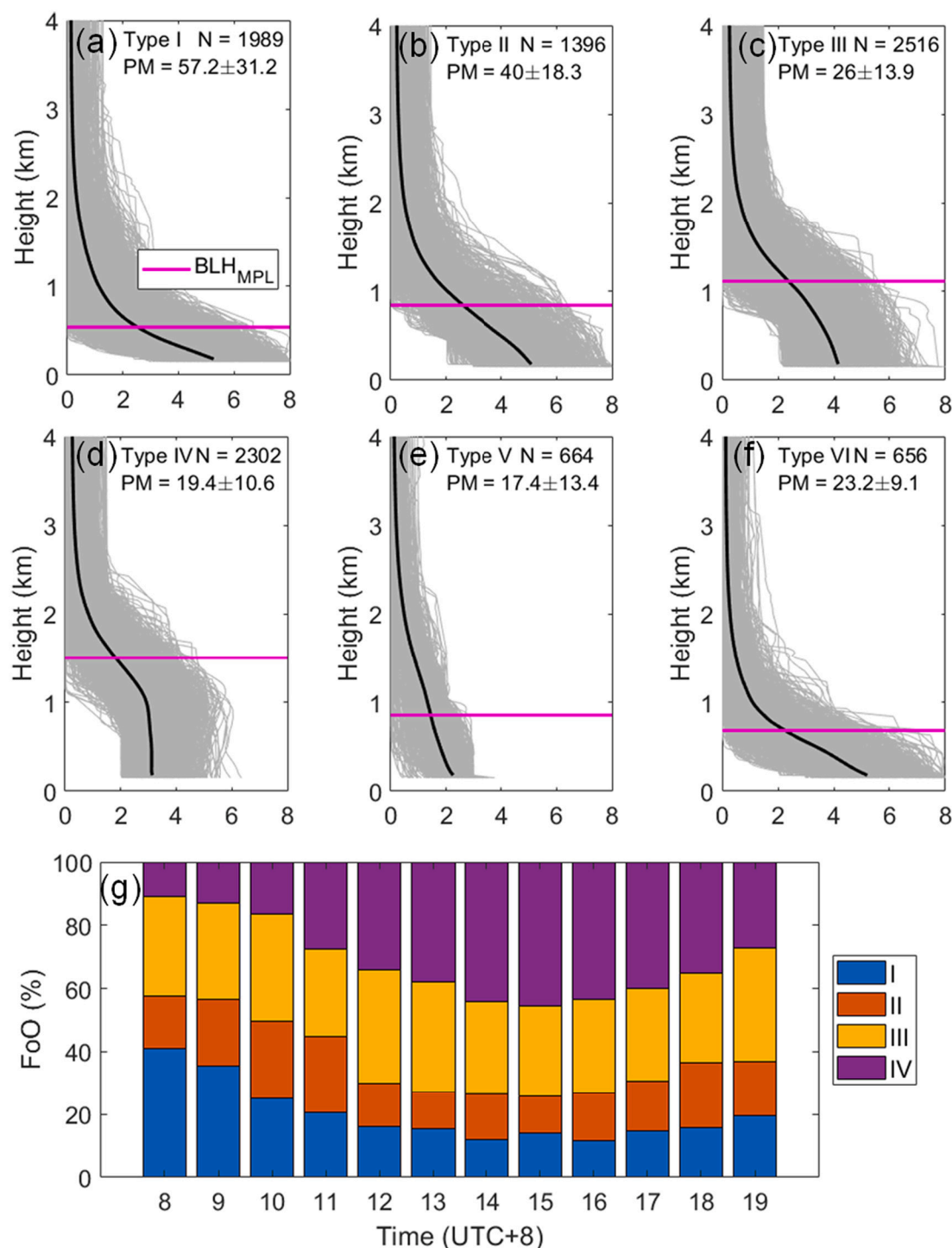
$$\text{HLH} = \text{BLH}_{\text{MPL}}(\alpha + \beta) \quad (3)$$

Fig. 3(a) and 3(b) present a demonstration of diurnal boundary layer evolution with the calculated parameters. A non-uniformly mixed layer would have  $\alpha < 1$  and large  $\beta$ , as shown in panel (c), resulting in HLH being reasonably higher than  $\text{MPL}_{\text{BLH}}$ . A relatively well-mixed layer would have  $\alpha$  close to 1 and small  $\beta$ , as shown in panel (d), resulting in an HLH resembling  $\text{BLH}_{\text{MPL}}$ .

### 2.3. Autoencoder-based deep residual network

An ADRN is trained for AOD imputation and  $\text{PM}_{2.5}$  estimation with HLH. The ADRN serves two main purposes in this work. First, it conducts spatial-temporal imputation of the MAIAC AOD to fill the missing data caused by gaps between satellite overpasses and cloud contamination. Second, it uses the hourly AOD from the output, HLH, and other ancillary information to predict the  $\text{PM}_{2.5}$  concentration. The framework is given in Fig. 2b, developed based on the python package for bagging deep residual neural networks (<https://pypi.org/project/baggingnet/>). The package contributor has used this residual network for the gap-filling of MAIAC AOD and AOD- $\text{PM}_{2.5}$  estimation, which achieved better performance than a regular neural network (Li, 2020a, 2020b; Li et al., 2020a). Only essential features of the ADRN are reiterated briefly here, and more details are provided in the referenced work.

The ADRN used in this work has a symmetrical structure from encoding layers, where complex and non-linear information from AOD and other input variables are recognized, to decoding layers, where the recognized patterns are used to reconstruct the original input data with the new features. It includes residual connection to provide shortcuts



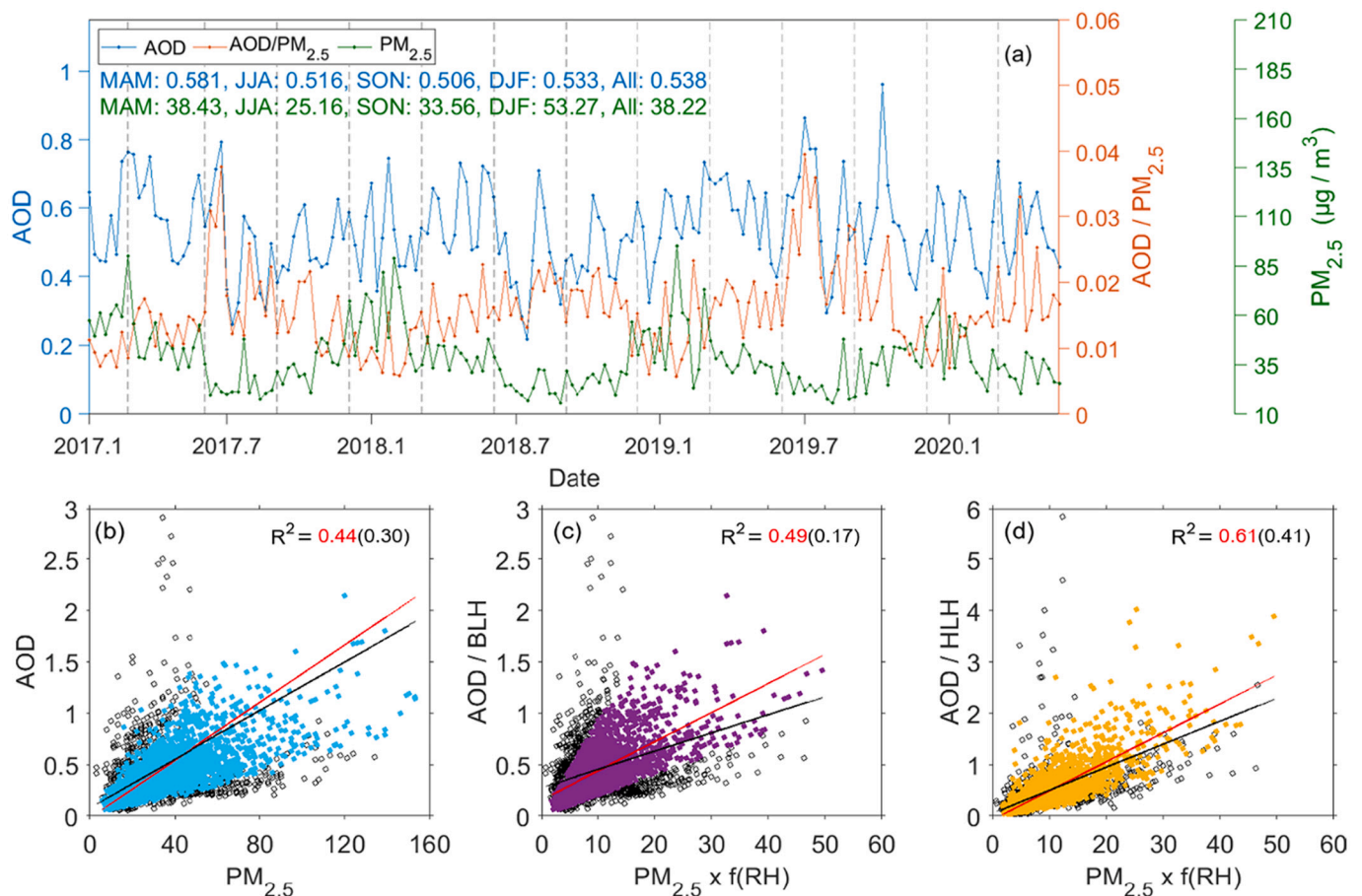
**Fig. 4.** Classification of MPL profiles and their frequency of occurrence. Panel (a-f) show 6 types of boundary layers, (a-d) are 4 major types classified by aerosol mixing condition, while (e-f) are 2 special types identified by mixing condition and ancillary information. The total number of profiles ( $N$ ) and average  $PM_{2.5}$  are given for each type. For 4 major types, frequency of occurrence through the day is present in panel (g).

among the layers to improve the efficient back-propagation of errors, which reduces the saturation and degradation of accuracy with increased hidden layers (Srivastava et al., 2015).

ADRN for AOD imputation contains AOD from the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al., 2017) and 13 other input variables, including temperature at 2 m above the surface, wind speed at 10 m above the surface, RH at  $10^3$  hPa above the surface, ground elevation, location, date, hour, and BLH. BLH from the numerical weather prediction (NWP)

model ( $BLH_{model}$ ) is used here for spatial variation. The spatiotemporal imputation is on weekly basis; the weeks before and after are also used in an index  $(-1, 0, 1)$ , where 0 is the current week.

ADRN for AOD- $PM_{2.5}$  estimation contains 2 additional inputs: interpolated AOD and ozone. HLH is used instead of  $BLH_{model}$ . Out of the total samples, 90% are used for training and 10% for testing. Bootstrap aggregation (bagging) is adopted to improve the accuracy and stability of the residual network by separately training multiple models using partial inputs and aggregating the results from each model to get the



**Fig. 5.** (a) Time-series of ground-based AOD measured by CE-318 and collocated PM<sub>2.5</sub> concentration from January 2017 to May 2020; (b) correlation between AOD and PM<sub>2.5</sub> concentration, in blue circles; (c) correlation between AOD/BLH and RH corrected PM<sub>2.5</sub> concentration, in purple circles; (d) correlation between AOD/HLH and RH corrected PM<sub>2.5</sub> concentration, in yellow circles. For (c-d), PM<sub>2.5</sub> concentration is corrected for hygroscopic growth  $f(\text{RH})$ . Solid circles account for 90% of all measurements (empty circles) used to calculate the  $R^2$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

best outcome (Gonzalez et al., 2020). Ten models are trained, each using 50% of the training data, and bagged for the final results. Besides validation based on the test dataset, we also compared the ADRN results with other machine learning models, including multilayer perceptron (MLP) feed-forward neural network (FFNN), support vector regression (SVR), and random forest (RF). Details of ADRN training and a brief introduction of each method are given in the supplement.

### 3. Results

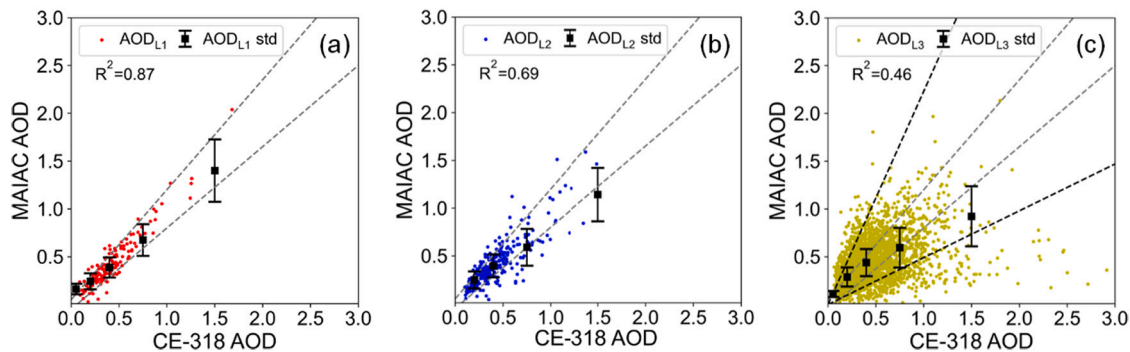
The results and analysis are presented in four subsections. The first one summarizes the diurnal evolution of HLH and related variables inferred from the MPL profile. The second subsection analyzes the AOD-PM<sub>2.5</sub> relationship after normalizing with BLH or HLH, respectively, based on measurements at the MPL station. The third one diagnoses the quality of AOD imputation through comparison with CE-318 observations and uses the imputed AOD for PM<sub>2.5</sub> estimation on the regional scale with HLH adaption through ADRN. The last subsection discusses the improved performance of ADRN and other models with HLH.

#### 3.1. Measurements of the atmospheric boundary layer using MPL

The retrieved boundary layer condition has a clear diurnal cycle, depending on weather, which suggests that correction for the non-uniform boundary layer is necessary in many cases. The background of local weather patterns in Hangzhou is present in the supplements.

To examine the variation of boundary layer and its influence on the estimation of ground-level PM<sub>2.5</sub> from space, we categorize a total of 9523 profiles into 6 types, mainly based on the average gradient from surface to BLH<sub>MPL</sub>. As shown in Fig. 4a-d, Type I-IV profiles indicate transition from the nocturnal stable boundary layer to the daytime convective boundary layer or vice versa. The frequency of occurrence (FoO, Fig. 4g) shows that Type I occurs most frequently during 8:00–10:00 (32.2%, 30.3%, and 22.0% of the total profiles, respectively). Type II occurs most frequently at 10:00–11:00 after the peak of Type I, and at 18:00 (32.2%, 30.3%, and 22.0%, respectively). Type III occur later at 12:00–13:00 and 19:00 (33.3%, 32.0%, and 31.3%, respectively). Type IV, represents the typical convective mixed boundary layer, which occurs most frequently from 14:00–16:00 (38.9%, 40.2%, 37.7%). The 4 major types of profiles identified depict the variation of boundary layer structures during the daytime due to the cease and return of solar heating. As the nocturnal stable boundary layer is stirred by solar heating of the surface, the BLH rises and becomes more uniform due to the convection of warm air. At the same time, aerosol particles near the surface mix with the surrounding air and the aloft residual layer from the previous day. This is confirmed by the frequency of occurrence that the non-uniform boundary layer (Type I) occurs more frequently in the morning with a large surface PM<sub>2.5</sub> concentration, while the well-mixed boundary layer (Type IV) occurs more frequently in the early afternoon with about a third of the average PM<sub>2.5</sub> in the morning.

From Type I to IV, the surface PM<sub>2.5</sub> (from nearby station) decreases



**Fig. 6.** Validation of AOD imputation results against CE-318 measurements at 3 levels: (a) actual observed ( $AOD_{L1}$ ), (b) interpolated when  $<20\%$  satellite pixels were missing ( $AOD_{L2}$ ) in the study region, and (c) interpolated when  $>20\%$  pixels were missing ( $AOD_{L3}$ ). The dotted gray lines mark the uncertainty of CE-318 measurements. The dotted black lines in panel (c) indicate the margin of 80% of  $AOD_{L3}$  used in the study.

as  $BLH_{MPL}$  increases,  $\alpha$  approaches 1 in general, and  $\beta$  decreases significantly from 0.92 to 0.18 on average. These trends suggest that aerosol particles become well-mixed and confined within the boundary layer, resulting in low  $PM_{2.5}$  concentration at the surface. Type V and VI (Fig. 4e-f) occur on special days and thus are not included in the daily transition. Type V profiles are measured on clean days. The average  $PM_{2.5}$  concentration at the surface is low ( $<20 \mu\text{g}/\text{m}^3$ ) even though no convection is observed. Type VI is identified by very strong signals near-surface, contradicting its low  $PM_{2.5}$  concentration ( $23.2 \pm 9.1 \mu\text{g}/\text{m}^3$ ). The intensity of the signal decreases sharply to a clean value below 0.5 km. Combined with the extremely high environmental humidity ( $RH = 89.8\%$ ), we conclude that Type VI represents conditions that aerosols are scavenged due to ground fog.

The analysis of MPL profiles confirms that BLH has a significant impact on surface  $PM_{2.5}$  concentration. Since the well-mixed (Type IV) or the relatively well-mixed (Type III) profiles only account for 55.2% of the total, correction for the non-uniform boundary layer is necessary for more than half of cases.

### 3.2. Correlation analysis of AOD/HLH- $PM_{2.5}$ at the MPL station

Following the variation of boundary layer height throughout the day, we use ground-based measurements to validate that AOD- $PM_{2.5}$  estimation can be further improved by normalizing AOD with HLH (AOD/HLH) over the traditional AOD/BLH normalization. Linear dependence is analyzed through the square of the Pearson correlation coefficient ( $R^2$ ) (Devore and Berk, 2012).

Fig. 5 illustrates the overall correlation between AOD and  $PM_{2.5}$ . Panel 4(a) presents the time-series variation of AOD and  $PM_{2.5}$  concentrations (plotted as weekly mean) from 2017 to 2020. There is no clear correlation observed between AOD and  $PM_{2.5}$  directly, because AOD shows little variation among different seasons (0.51–0.58), while  $PM_{2.5}$  concentration shows a clear maximum in winter ( $53.27 \mu\text{g}/\text{m}^3$ ) and minimum in summer ( $25.16 \mu\text{g}/\text{m}^3$ ). After normalizing with BLH (Fig. 5c) or HLH (Fig. 5d), 2673 sets of coincident measurements of ground-based AOD and  $PM_{2.5}$  concentration show weak to moderate correlations.

For the full dataset, normalizing AOD with HLH increases the correlation to  $R^2 = 0.41$  compared to  $R^2 = 0.17$  for normalizing with BLH. Since a small fraction of measurements with high AOD/ $PM_{2.5}$  ratios might be affected by rare cases of regional transport and other measurement errors might occur, we remove 10% of total measurements with the highest or lowest AOD/ $PM_{2.5}$  ratio. For the remaining 90% of measurements, AOD/BLH has a moderate correlation with  $PM_{2.5}$  concentration ( $R^2 = 0.49$ ), which increased to a relatively high correlation of  $R^2 = 0.61$  if normalized with HLH. Both normalization methods of AOD yield better results than the unnormalized dataset, though the increase with BLH is very limited. For this comparison, the effect of

hygroscopic growth is considered when assessing the AOD- $PM_{2.5}$  correlation. The hygroscopic growth factor  $f(RH)$  is defined as:

$$f(RH) = \frac{1}{(1-RH)^\gamma} \quad (4)$$

where  $\gamma$  is the coefficient of growth. The  $\gamma$  is set to 0.35, adapted from a campaign conducted at locations with similar aerosol types (Chu et al., 2015).

The increase in correlation agreed with previous studies showing aerosol vertical distribution is a key factor in the relationship between AOD and  $PM_{2.5}$  (He et al., 2008; Boyouk et al., 2010). By taking into account aerosol mixing in the boundary layer and residual above, HLH calculated from the lidar profile helps correlate column AOD with surface aerosol mass concentration. Some relatively poor results in the simple relationship between AOD and  $PM_{2.5}$  due to non-uniform distribution demonstrate significant improvements. Such results validate the potential of HLH for improving  $PM_{2.5}$  estimation, but further employment in the regional  $PM_{2.5}$  estimation requires sufficient AOD measurements in the studied area. In this study, regional AOD data are obtained through satellite measurements and ADRN-based imputation technique.

### 3.3. Assessment of AOD imputation and regional AOD- $PM_{2.5}$ estimation

Following the correlation analysis of simultaneous measurements at the MPL station, MAIAC AOD with hourly imputation is used to test the validity of using HLH from a single lidar to improve the  $PM_{2.5}$  estimation in the study area. As stated, the imputation of AOD follows previous work using ADRN (Li, 2020a, 2020b; Li et al., 2020a). The imputed AOD is categorized into 3 levels:  $AOD_{L1}$  is the actual observation;  $AOD_{L2}$  is interpolated AOD when the proportion of missing AOD is  $<20\%$ ;  $AOD_{L3}$  is interpolated AOD when the proportion of missing data is  $>20\%$ . Each level of results is validated against CE-318 measurements (Fig. 6).

The comparison shows that the ADRN imputation results agree well with coincident measurements from ground-based CE-318 sun-photometer for  $AOD_{L1}$  (Fig. 6a) and  $AOD_{L2}$  (Fig. 6b). The segmented mean and standard deviation (std) of observed MAIAC AOD and interpolated results are given in Table S2.  $AOD_{L1}$  data correlates strongly ( $R^2 = 0.86$ ,  $RMSE = 0.10$ ) with the ground-based AOD, and mean values always fall in the uncertainty of CE-318 measurements.  $AOD_{L2}$  data also correlates well with ground-based AOD ( $R^2 = 0.69$ ,  $RMSE = 0.14$ ), but tend to underestimate in high aerosol loading cases. Over the 3.5-yr study period, the observed MAIAC AOD ( $AOD_{L1}$ ) accounts for 4.96% of spatial-temporal coverage, and the partly interpolated  $AOD_{L2}$  accounts for 8.3%.

$AOD_{L3}$  is interpolated with a high proportion of missing values due to cloud cover; therefore, it is difficult to fill the gaps and consequently,  $AOD_{L3}$  has a relatively wide span. We limited the overall coverage to

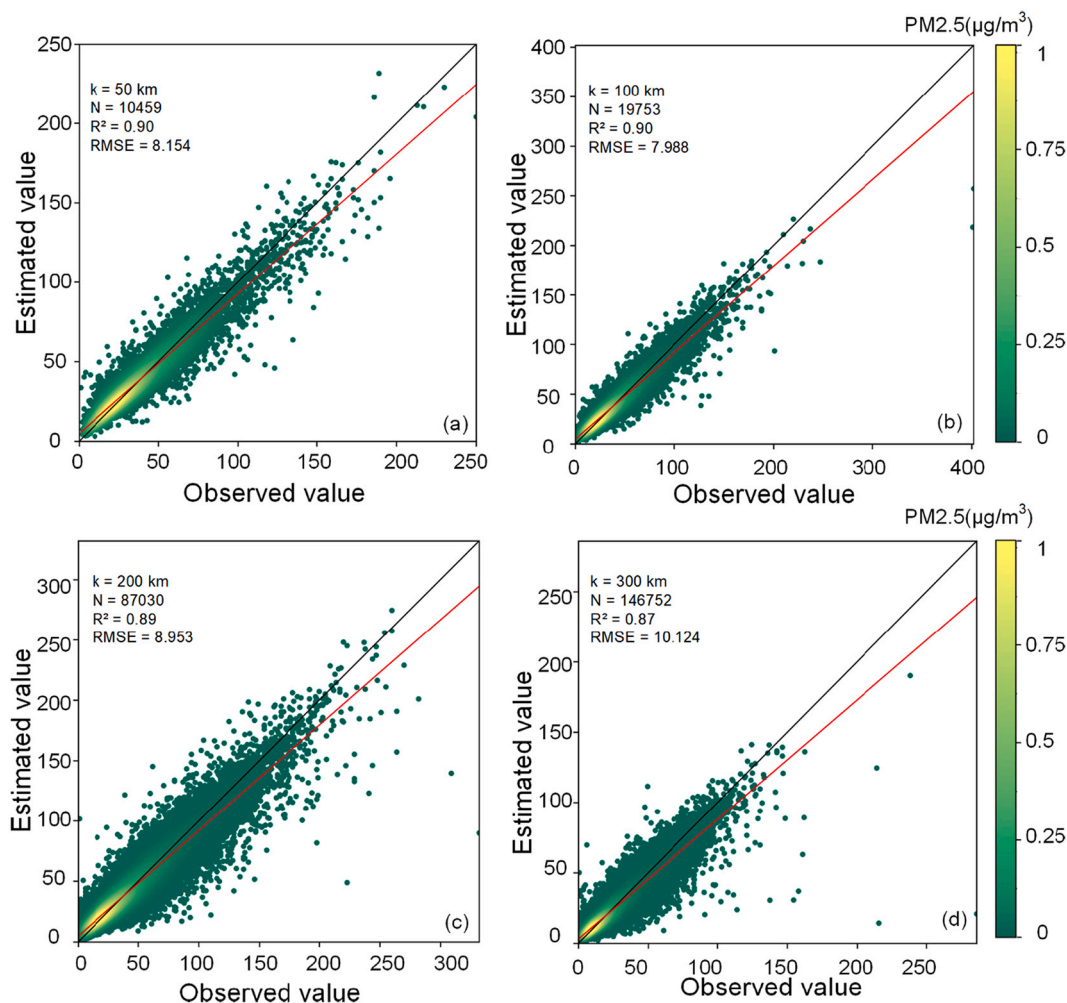


Fig. 7. The scatter plot of ADRN models with point density using data within (a) 50 km, (b) 100 km, (c) 200 km, and (d) 300 km of the MPL station. The black line is the 1-to-1 ratio line, and the red line is separate fitting lines between estimated  $PM_{2.5}$  concentration and observed values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Estimated  $PM_{2.5}$  from 4 machine learning models compared to observed values at AQI stations, using BLH or HLH, respectively.

With BLH			2017		2018		2019		2020	
Method	$R^2$	Mean/bias	DJF	Other	DJF	Other	DJF	Other	DJF	Other
AQI		<b>38.68</b>	58.05	39.73	61.16	31.97	62.46	32.59	45.92	30.88
MLP	<b>0.34</b>	-0.14	-6.37	0.98	-4.45	1.79	-4.10	0.47	-4.56	2.43
SVR	<b>0.36</b>	-2.40	-8.84	-1.94	-11.43	-1.96	-8.88	-1.56	-7.48	-0.93
RF	<b>0.72</b>	<b>0.06</b>	-1.34	-0.04	-0.21	0.19	-0.37	-0.02	-0.20	0.14
ADRN	<b>0.80</b>	<b>0.12</b>	-0.98	-0.05	-0.26	0.11	-0.08	0.03	-0.01	0.28
With HLH			2017		2018		2019		2020	
Method	$R^2$	Mean/bias	DJF	Other	DJF	Other	DJF	Other	DJF	Other
AQI		<b>38.68</b>	58.05	39.73	61.16	31.97	62.46	32.59	45.92	30.88
MLP	<b>0.44</b>	-0.11	-0.84	0.91	-4.66	0.03	-3.25	0.43	-1.82	0.94
SVR	<b>0.40</b>	-2.22	-4.40	-1.95	-7.06	-1.85	-5.53	-1.21	-4.81	-1.10
RF	<b>0.83</b>	-0.06	-0.28	0.15	-1.23	0.16	-0.77	0.18	0.09	0.42
ADRN	<b>0.87</b>	-0.01	-0.18	0.20	-0.08	0.15	0.39	0.13	-0.07	0.14

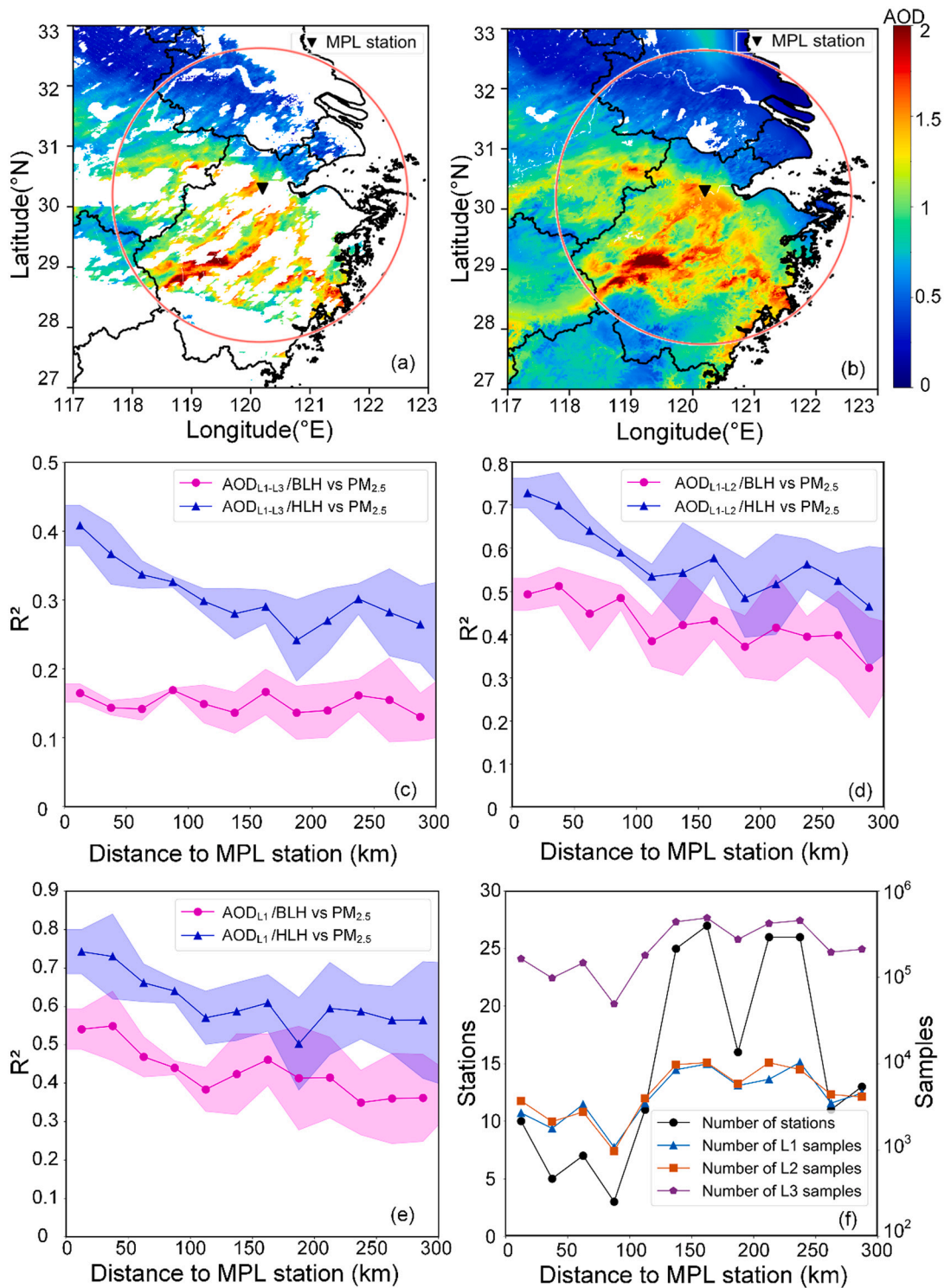
80% by removing days subject to large errors. The remaining data show moderate correlation ( $R^2 = 0.46$ ,  $RMSE = 0.19$ ) with ground-based AOD.

### 3.4. Improvement in regional $PM_{2.5}$ estimation using HLH

Using interpolated AOD at all three levels, ADRN models for AOD/

HLH- $PM_{2.5}$  estimation are trained with different distances ( $k$ ) to the MPL station, and the performance is first examined through an independent test using 10% of samples. Fig. 7 presents the scatter plots with point density of observed  $PM_{2.5}$  concentration and estimated values at  $k = 50, 100, 200,$  and  $300$  km, respectively. The test  $R^2$  of trained models decreases only slightly as the range increases; models using data within 100 km of the MPL station have the highest  $R^2 = 0.90$  ( $RMSE = 7.99 \mu\text{g}/$

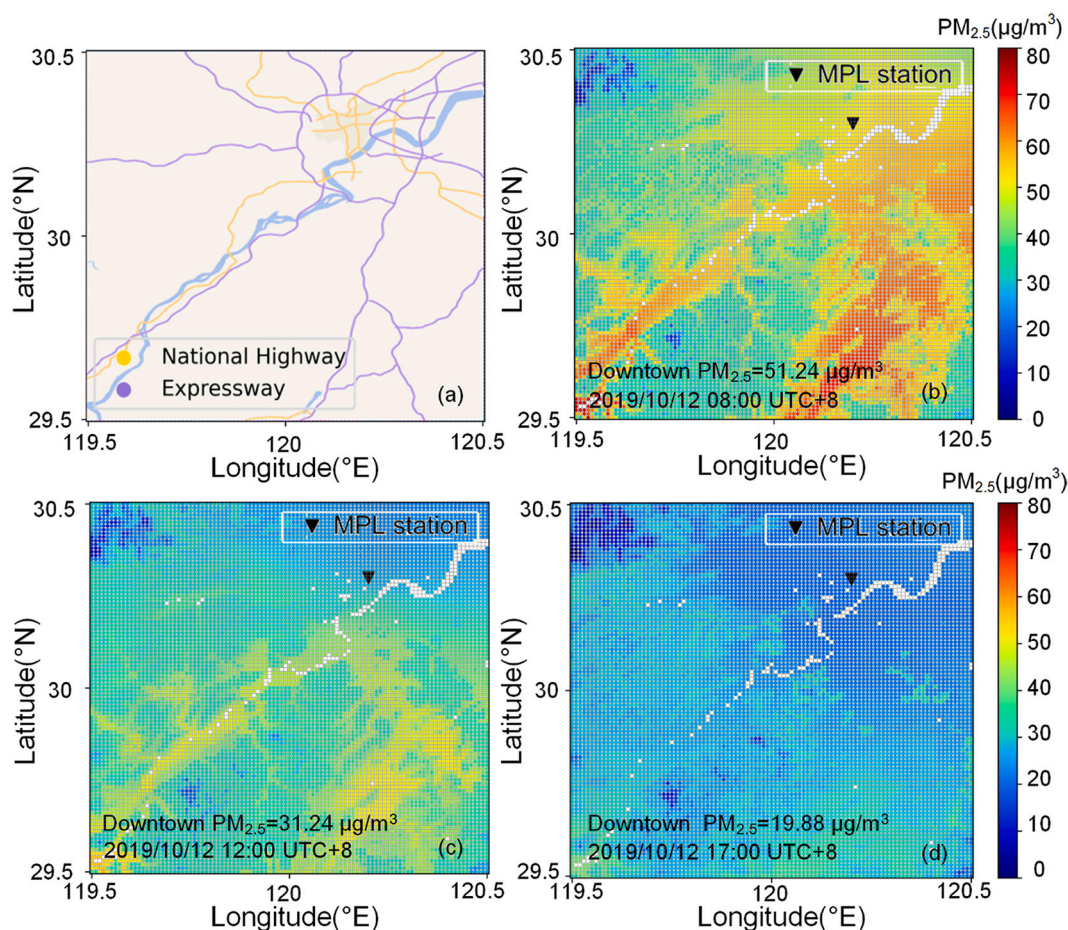




**Fig. 8.** AOD and PM<sub>2.5</sub> concentration from ADRN. Panel (a) shows observed MAIAC AOD (AOD<sub>L1</sub>) while panel (b) shows observed and interpolated AOD (AOD<sub>L2</sub> since >20% pixel is available) demonstrated by data on January 14th, 2017. The red circle marks the boundary of the study area. (c) Correlation between all available AOD (AOD<sub>L1-L3</sub>) and PM<sub>2.5</sub> concentration measured at AQI stations, with and without normalizing by HLH, arranged by distance to MPL station. (d) Same as (c) but using observed and partially interpolated AOD (AOD<sub>L1-L2</sub>) only. (e) Same as (c) but using observed AOD (AOD<sub>L1</sub>) only. (f) The number of AQI stations and collocated samples averaged by distance to the MPL station. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

m<sup>3</sup>). Since HLH is inferred from MPL measurements and assumed to be spatially uniform in the range of the trained model,  $k$  above the feasible value might negatively impact the validity of the model. Yet, the ADRN achieves competitive performance in PM<sub>2.5</sub> estimation up to 300 km, with  $R^2 = 0.87$  and  $RMSE = 10.14 \mu\text{g}/\text{m}^3$ .

To prove that using HLH instead of BLH has a positive effect on the model results, we performed a comprehensive comparison among PM<sub>2.5</sub> observed at stations and estimated PM<sub>2.5</sub> using machine learning models including MLP, SVR, RF, and ADRN. The statistics are given in Table.1. The results in form of seasonal bar plots are included in Fig.S2. The



**Fig. 9.** A case study of estimated hourly  $PM_{2.5}$  concentration on October 12th, 2019. Panel (a) depicts the major transportation network surrounding Hangzhou. Panels (b-d) show spatial distributions of estimated hourly  $PM_{2.5}$  concentration with 1-km-resolution at 8:00, 12:00 and 17:00 local time, respectively.

comparison shows that all models are improved using HLH instead of BLH. The improvement in correlation is most prominent in the RF model, and the reduction of bias is most prominent in the ADRN model. Note that there might be some sampling bias when comparing measurements at AQI stations to the model results, since the stations are unevenly distributed, but the influence is minor in the current study (see supplement).

#### 4. Discussion

Using HLH to improve regional  $PM_{2.5}$  estimation on an hourly basis has important applications in extending spatial and temporal coverage of  $PM_{2.5}$  in poorly monitored areas. Fig. 8 demonstrates the feasibility of using HLH from a single lidar to improve the accuracy of  $PM_{2.5}$  estimation in the surrounding area, with the aid of AOD imputation. Fig. 8 (a) presents the spatial availability of MAIAC AOD on January 14th, 2017, and Fig. 8(b) is the interpolated AOD using ADRN. The missing area is due to the water surface, which is excluded from interpolation because MAIAC AOD is subject to large uncertainties. Fig. 8(c) presents the improvement in estimated  $PM_{2.5}$  for all AOD, while Fig. 8(d-e) only uses  $AOD_{L1-L2}$  and  $AOD_{L1}$  only for higher accuracy. The results are averaged by the distance from each AQI station to the MPL station, which roughly shows an inverse relationship to the improvement of AOD/HLH- $PM_{2.5}$  estimation. The number of AQI stations and collocated samples are shown in Fig. 8(f). For the full dataset, the correlations between  $AOD_{L1-L3}$  and  $PM_{2.5}$  concentration at individual stations improve from  $R^2 = 0.13-0.17$  to  $0.26-0.41$  after normalizing AOD with HLH instead of BLH. For  $AOD_{L1-L2}$ , the correlation improves from  $R^2 =$

$0.32-0.51$  to  $0.47-0.73$ . For  $AOD_{L1}$ , the correlation improves from  $R^2 = 0.36-0.54$  to  $0.56-0.74$ . Comparison between results using  $AOD_{L1-L2}$  and  $AOD_{L1}$  only shows that a small increase in data coverage ( $AOD_{L1}$  accounts for 4.96% of total and  $AOD_{L2}$  accounts for 8.3%) doesn't significantly impact the model performance. When the distance is larger than 125 km, the improvement still exists, but the trend becomes unclear, indicating that a single lidar is effective to improve the accuracy of  $PM_{2.5}$  estimates up to this range.

After accommodating non-uniform boundary layer conditions, accurate estimates of hourly  $PM_{2.5}$  can have important applications for social issues. As shown in Fig. 9, a case study at 8:00, 12:00, and 17:00, on October 12th, 2019 reveals significant changes during daytime surrounding the city. Fig. 9a shows high  $PM_{2.5}$  concentration ( $51.24 \mu\text{g}/\text{m}^3$ ) in the downtown region (encircled by expressway) accumulated from the previous day and along with the major transportation network due to morning rush hour. On average, the north wind blows from morning to noon at 7.78 mph, and the  $PM_{2.5}$  concentration in the downtown area is reduced to  $31.24 \mu\text{g}/\text{m}^3$  by noon. The wind keeps blowing in the afternoon, and the downtown region shows very low  $PM_{2.5}$  concentration ( $19.88 \mu\text{g}/\text{m}^3$ ) at 17:00 before the evening traffic hour, which has decreased by  $\sim 60\%$  compared with that at 8:00. Large variations in  $PM_{2.5}$  concentrations suggest that hourly air quality monitoring, rather than daily averages, is important for correctly grading air quality warnings.

Many studies have investigated the relation between  $PM_{2.5}$  exposure and human adult life expectancy around the world. The ADRN results show that the highest  $PM_{2.5}$  concentration occurs at 9:00 for winter and 8:00 for other seasons (Table S3). The difference between the morning

concentration peak and average daytime value is  $5.35 \mu\text{g}/\text{m}^3$  (the largest difference is  $7.62 \mu\text{g}/\text{m}^3$  in winter). Based on equations provided in the  $\text{PM}_{2.5}$ -life expectancy studies (Krstic, 2012; Correia et al., 2013; Stieb et al., 2015; Chen et al., 2019), this value could equate to 0.19–2.57 (0.27–2.67 in winter) years of life expectancy change (see equations provided in the supplement).

## 5. Conclusion

In order to accurately estimate hourly  $\text{PM}_{2.5}$  concentration, we proposed the CAPTA method combining the advantage of active and passive remote sensing data, and calculated an HLH parameter from MPL profiles to take the non-uniform boundary layer and the residual layer above into account. Long-term lidar observation shows that due to non-uniform boundary layers, about half AOD-based  $\text{PM}_{2.5}$  estimates in this study can be improved by replacing BLH with HLH as the scaling factor. For 90% of measurements at the MPL station, normalizing with HLH improves the correlation between AOD and  $\text{PM}_{2.5}$  concentration by 20% ( $R^2$  increased from 0.49 to 0.61). On the regional scale, employing HLH to the observed MAIAC AOD and interpolated AOD using ADRN also leads to improvement in correlation with  $\text{PM}_{2.5}$  concentration measured at each AQI station. For observed and partially interpolated AOD ( $\text{AOD}_{L1}$  and  $\text{AOD}_{L2}$ , respectively), the correlation improves from  $R^2 = 0.32$ – $0.51$  to  $0.47$ – $0.73$ . For all available AOD, the correlation improves from  $R^2 = 0.13$ – $0.17$  to  $0.26$ – $0.41$ . The improvements decrease as the distance to MPL increases. Such a trend is most clear in the first 125 km, suggesting that a single lidar is feasible to improve the accuracy of  $\text{PM}_{2.5}$  estimates up this range.

Trained ADRN models using HLH as one of the input variables demonstrate high performance in the independent test. The test  $R^2$  is 0.90 (RMSE =  $7.99 \mu\text{g}/\text{m}^3$ ) within 100 km of MPL and 0.87 within 300 km (RMSE =  $10.14 \mu\text{g}/\text{m}^3$ ). Estimated  $\text{PM}_{2.5}$  concentration at high spatial-temporal resolutions is consistent with daytime traffic, one of the major sources of urban particulates.  $\text{PM}_{2.5}$  estimates from other machine learning models also show improvement using HLH instead of BLH. The high spatial-temporal variation of  $\text{PM}_{2.5}$  indicates that hourly monitoring is the key to social health. The average difference between morning and daily  $\text{PM}_{2.5}$  concentration could equate to 0.19–2.57 years of life expectancy change.

This work demonstrates that the CAPTA method is useful for accommodating non-uniform boundary layer conditions and estimating  $\text{PM}_{2.5}$  accurately. Ground-based lidar has been shown to be a powerful tool for investigating the local pattern of aerosol vertical distribution and provides correction coefficients for the  $\text{PM}_{2.5}$  calculation. Yet, the limited availability of lidar makes it challenging to be employed in routine  $\text{PM}_{2.5}$  monitoring. The developing deep learning approaches provide a helpful technique for combining profiles from ground-based lidar and AOD from space-borne instruments on the regional scale. Furthermore, we look forward to launching more space-borne lidars or establishing a ground-based lidar network to expand the coverage of lidar profiles.

## Author contributions

Conceptualization, D.L.\* and S.C.; methodology, S.C., L.W., B.T. and C.X.; data curation, B.Q., Y.X., H.Z., and D.L.\*; writing-original draft, S.C., Y.S., and B.Z.; writing-review & editing, L.M.R., J.W., J.G., F.M., B.Q., Y.X., H.Z., D.L., and Z.H.; funding acquisition, D.L.\*.

## Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Bing Qi reports financial support was provided by Hangzhou Meteorological Bureau. Dong Liu has patent #CN113344149B licensed to Zhejiang University.

## Data availability

Data will be made available on request.

## Acknowledgments

The study was supported by the National Key Research and Development Program of China (2016YFC1400900); Scientific Research Projects of Zhejiang Administration for Market Regulation(20200103); Excellent Young Scientist Program of Zhejiang Provincial Natural Science Foundation of China (LR19D050001); Agriculture and Social Development Research Project of Hangzhou (20201203B155); the Key Project of Science and Technology Plan of Zhejiang Meteorological Bureau (2021ZD13) and Zhejiang Provincial Basic Public Welfare Research Project (LGF22D050004).

We thank the following people or organizations for providing the data and package used in this paper and list the resources here. ECWMF data are available at <https://cds.climate.copernicus.eu>. MERRA-2 data are available at <https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI>. MODIS data are downloaded from the Land Processes Distributed Active Archive Center (LP DAAC) (<https://ladsweb.modaps.eosdis.nasa.gov>). Hourly  $\text{PM}_{2.5}$  concentrations are provided by the Ministry of Ecology and Environment of China (<http://106.37.208.233:20035>). Package for bagging of deep residual neural networks is provided by professor Lianfa Li (<https://pypi.org/project/baggingrnet/>).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2022.113224>.

## References

- Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K.S., Christensen, M., Daniau, A.L., Dufresne, J.L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J.M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D.T., Myhre, G., Mulmenstadt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S.E., Sourdval, O., Storelvmo, T., Toll, V., Winker, D., Stevens, B., 2020. Bounding global aerosol radiative forcing of climate change. *Rev. Geophys.* 58 (1).
- Berkoff, F.A., Welton, E.J., Campbell, J.R., Valencia, S., Spinhirne, J.D., Tsay, S.C., Holben, B.N., *ieee*, 2004. Observations of aerosols using the Micro-Pulse Lidar NETwork (MPLNET). In: *Igarss 2004: Ieee International Geoscience and Remote Sensing Symposium Proceedings, Vols 1–7: Science for Society: Exploring and Managing a Changing Planet*, pp. 2208–2211.
- Boyouk, N., Leon, J.-F., Delbarre, H., Podvin, T., Deroo, C., 2010. Impact of the mixing boundary layer on the relationship between  $\text{PM}_{2.5}$  and aerosol optical thickness. *Atmos. Environ.* 44 (2), 271–277.
- Chen, C.C., Chen, P.S., Yang, C.Y., 2019. Relationship between fine particulate air pollution exposure and human adult life expectancy in Taiwan. *J. Toxicol. Environ. Health-Part a-Curr. Issue* 82 (14), 826–832.
- Chen, X., de Leeuw, G., Arola, A., Liu, S., Liu, Y., Li, Z., Zhang, K., 2020. Joint retrieval of the aerosol fine mode fraction and optical depth using MODIS spectral reflectance over northern and eastern China: artificial neural network method. *Remote Sens. Environ.* 249.
- Chu, D.A., Tsai, T.-C., Chen, J.-P., Chang, S.-C., Jeng, Y.-J., Chiang, W.-L., Lin, N.-H., 2013. Interpreting aerosol lidar profiles to better estimate surface  $\text{PM}_{2.5}$  for columnar AOD measurements. *Atmos. Environ.* 79, 172–187.
- Chu, D.A., Ferrare, R., Szykman, J., Lewis, J., Scarino, A., Hains, J., Burton, S., Chen, G., Tsai, T., Hostetler, C., Hair, J., Holben, B., Crawford, J., 2015. Regional characteristics of the relationship between columnar AOD and surface  $\text{PM}_{2.5}$ : application of lidar aerosol extinction profiles over Baltimore-Washington Corridor during DISCOVER-AQ. *Atmos. Environ.* 101, 338–349.
- Cordoba-Jabonero, C., Sicard, M., Ansmann, A., del Aguila, A., Baars, H., 2018. Separation of the optical and mass features of particle components in different aerosol mixtures by using POLIPHON retrievals in synergy with continuous polarized Micro-Pulse Lidar (P-MPL) measurements. *Atmos. Meas. Tech.* 11 (8), 4775–4795.
- Correia, A.W., Pope III, C.A., Dockery, D.W., Wang, Y., Ezzati, M., Dominici, F., 2013. Effect of air pollution control on life expectancy in the United States an analysis of 545 US counties for the period from 2000 to 2007. *Epidemiology* 24 (1), 23–31.
- Devore, J.L., Berk, K.N., 2012. *Modern Mathematical Statistics with Applications*. Springer, New York; London.
- Emeis, S., Schaefer, K., Muenkel, C., 2008. Surface-based remote sensing of the mixing-layer height - a review. *Meteorol. Z.* 17 (5), 621–630.

- Geiss, A., Wiegner, M., Bonn, B., Schaefer, K., Forkel, R., von Schneidmesser, E., Muenkel, C., Chan, K.L., Nothard, R., 2017. Mixing layer height as an indicator for urban air quality? *Atmos. Meas. Tech.* 10 (8), 2969–2988.
- Gelaro, R., McCarty, W., Suarez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* 30 (14), 5419–5454.
- Geng, G., Zhang, Q., Tong, D., Li, M., Zheng, Y., Wang, S., He, K., 2017. Chemical composition of ambient PM<sub>2.5</sub> over China and relationship to precursor emissions during 2005–2012. *Atmos. Chem. Phys.* 17 (14), 9187–9203.
- Gonzalez, S., Garcia, S., Del Ser, J., Rokach, L., Herrera, F., 2020. A practical tutorial on bagging and boosting based ensembles for machine learning: algorithms, software tools, performance study, practical perspectives and opportunities. *Inform. Fusion* 64, 205–237.
- Gupta, S.K., Elumalai, S.P., 2019. Dependence of urban air pollutants on morning/evening peak hours and seasons. *Arch. Environ. Contam. Toxicol.* 76 (4), 572–590.
- Hasheminassab, S., Pakbin, P., Delfino, R.J., Schauer, J.J., Sioutas, C., 2014. Diurnal and seasonal trends in the apparent density of ambient fine and coarse particles in Los Angeles. *Environ. Pollut.* 187, 1–9.
- He, Q.S., Li, C.C., Mao, J.T., Lau, A.K.H., Chu, D.A., 2008. Analysis of aerosol vertical distribution and variability in Hong Kong. *J. Geophys. Res.-Atmos.* 113 (D14).
- Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Al-Hamdan, M.Z., Crosson, W.L., Estes Jr., M.G., Estes, S.M., Quattrochi, D.A., Puttaswamy, S.J., Liu, Y., 2014. Estimating ground-level PM<sub>2.5</sub> concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sens. Environ.* 140, 220–232.
- Huang, J., Hsu, N.C., Tsay, S.-C., Holben, B.N., Welton, E.J., Smirnov, A., Jeong, M.-J., Hansell, R.A., Berkoff, T.A., Liu, Z., Liu, G.-R., Campbell, J.R., Liew, S.C., Barnes, J. E., 2012. Evaluations of cirrus contamination and screening in ground aerosol observations using collocated lidar systems. *J. Geophys. Res.-Atmos.* 117.
- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., Liu, Y., 2018. Predicting monthly high-resolution PM<sub>2.5</sub> concentrations with random forest model in the North China Plain. *Environ. Pollut.* 242, 675–683.
- Huang, X., Ding, A., Wang, Z., Ding, K., Gao, J., Chai, F., Fu, C., 2020. Amplified transboundary transport of haze by aerosol-boundary layer interaction in China. *Nat. Geosci.* 13 (6), 428–4.
- Krstic, G., 2012. A reanalysis of fine particulate matter air pollution versus life expectancy in the United States. *J. Air Waste Manage. Assoc.* 62 (9).
- Levy, R.C., Mattoo, S., Munchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., Hsu, N.C., 2013. The collection 6 MODIS aerosol products over land and ocean. *Atmos. Meas. Tech.* 6 (11), 2989–3034.
- Lewis, J., De Young, R., Ferrare, R., Chu, D.A., 2010. Comparison of summer and winter California central valley aerosol distributions from lidar and MODIS measurements. *Atmos. Environ.* 44 (35), 4510–4520.
- Li, L., 2020a. A robust deep learning approach for spatiotemporal estimation of satellite AOD and PM<sub>2.5</sub>. *Remote Sens.* 12 (2).
- Li, L.F., 2020b. A robust deep learning approach for spatiotemporal estimation of satellite AOD and PM<sub>2.5</sub>. *Remote Sens.* 12 (2).
- Li, Z.Q., Guo, J.P., Ding, A.J., Liao, H., Liu, J.J., Sun, Y.L., Wang, T.J., Xue, H.W., Zhang, H.S., Zhu, B., 2017. Aerosol and boundary-layer interactions and impact on air quality. *Nat. Sci. Rev.* 4 (6), 810–833.
- Li, X.-B., Wang, D.-S., Lu, Q.-C., Peng, Z.-R., Wang, Z.-Y., 2018. Investigating vertical distribution patterns of lower tropospheric PM<sub>2.5</sub> using unmanned aerial vehicle measurements. *Atmos. Environ.* 173, 62–71.
- Li, L., Franklin, M., Girguis, M., Lurmann, F., Wu, J., Pavlovic, N., Breton, C., Gilliland, F., Habre, R., 2020a. Spatiotemporal imputation of MAIAC AOD using deep learning with downsampling. *Remote Sens. Environ.* 237.
- Li, H., Yang, Y., Wang, H., Li, B., Wang, P., Li, J., Liao, H., 2020b. Constructing a spatiotemporally coherent long-term PM<sub>2.5</sub> concentration dataset over China during 1980–2019 using a machine learning approach. *Sci. Total Environ.* 765, 144263.
- Liang, F., Xiao, Q., Huang, K., Yang, X., Liu, F., Li, J., Lu, X., Liu, Y., Gu, D., 2020. The 17-y spatiotemporal trend of PM<sub>2.5</sub> and its mortality burden in China. *Proc. Natl. Acad. Sci. U. S. A.* 117 (41), 25601–25608.
- Lin, C.Q., Li, Y., Yuan, Z.B., Lau, A.K.H., Li, C.C., Fung, J.C.H., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM<sub>2.5</sub>. *Remote Sens. Environ.* 156, 117–128.
- Liu, J., Wu, Q., Sui, X., Chen, Q., Gu, G., Wang, L., Li, S., 2021. Research progress in optical neural networks: theory, applications and developments. *PhotonIX* 2 (1), 5.
- Nakoudi, K., Giannakaki, E., Dandou, A., Tombrou, M., Komppula, M., 2019. Planetary boundary layer height by means of lidar and numerical simulations over New Delhi, India. *Atmos. Meas. Tech.* 12 (5), 2595–2610.
- Pal, S., Lopez, M., Schmidt, M., Ramonet, M., Gibert, F., Xueref-Remy, I., Ciais, P., 2015. Investigation of the atmospheric boundary layer depth variability and its impact on the Rn-222 concentration at a rural site in France. *J. Geophys. Res.-Atmos.* 120 (2), 623–643.
- Reid, J.S., Kuehn, R.E., Holz, R.E., Eloranta, E.W., Kaku, K.C., Kuang, S., Newchurch, M. J., Thompson, A.M., Trepte, C.R., Zhang, J., Atwood, S.A., Hand, J.L., Holben, B.N., Minnis, P., Posselt, D.J., 2017. Ground-based high spectral resolution Lidar observation of aerosol vertical distribution in the summertime Southeast United States. *J. Geophys. Res.-Atmos.* 122 (5), 2970–3004.
- Schneider, J., Eixmann, R., 2002. Three years of routine Raman lidar measurements of tropospheric aerosols: backscattering, extinction, and residual layer height. *Atmos. Chem. Phys.* 2, 313–323.
- Shen, X., Wang, N., Veselovskii, I., Xiao, D., Zhong, T., Liu, C., Zhang, K., Zhou, Y., Liu, D., 2020. Development of ZJU high-spectral-resolution lidar for aerosol and cloud: calibration of overlap function. *J. Quant. Spectrosc. Radiat. Transf.* 257.
- Song, W.Z., Jia, H.F., Huang, J.F., Zhang, Y.Y., 2014. A satellite-based geographically weighted regression model for regional PM<sub>2.5</sub> estimation over the Pearl River Delta region in China. *Remote Sens. Environ.* 154, 1–7.
- Sorek-Hamer, M., Strawa, A.W., Chaffield, R.B., Esswein, R., Cohen, A., Broday, D.M., 2013. Improved retrieval of PM<sub>2.5</sub> from satellite data products using non-linear methods. *Environ. Pollut.* 182, 417–423.
- Srivastava, R.K., Greff, K., Schmidhuber, J., 2015. Training very deep networks. *Adv. Neural Inf. Process. Syst.* 28 (Nips 2015) 28.
- Stieb, D.M., Judek, S., van Donkelaar, A., Martin, R.V., Brand, K., Shin, H.H., Burnett, R. T., Smith-Doiron, M.H., 2015. Estimated public health impacts of changes in concentrations of fine particle air pollution in Canada, 2000 to 2011. *Canadian Journal of Public Health-Revue Canadienne De Sante Publique* 106 (6), E362–E368.
- Sun, Y., Zeng, Q., Geng, B., Lin, X., Sude, B., Chen, L., 2019. Deep learning architecture for estimating hourly ground-level PM<sub>2.5</sub> using satellite remote sensing. *IEEE Geosci. Remote Sens. Lett.* 16 (9), 1343–1347.
- Tian, J., Chen, D., 2010. A semi-empirical model for predicting hourly ground-level fine particulate matter (PM<sub>2.5</sub>) concentration in southern Ontario from satellite remote sensing and ground-based meteorological measurements. *Remote Sens. Environ.* 114 (2), 221–229.
- Tsaknakis, G., Papayannis, A., Kokkalis, P., Amiridis, V., Kambezidis, H.D., Mamouri, R. E., Georgoussis, G., Avdikos, G., 2011. Inter-comparison of lidar and ceilometer retrievals for aerosol and planetary boundary layer profiling over Athens, Greece. *Atmos. Meas. Tech.* 4 (6), 1261–1273.
- van Donkelaar, A., Martin, R.V., Li, C., Burnett, R.T., 2019. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol.* 53 (5), 2595–2611.
- Wang, W., Zhao, S., Jiao, L., Taylor, M., Zhang, B., Xu, G., Hou, H., 2019. Estimation of PM<sub>2.5</sub> concentrations in China using a spatial back propagation neural network. *Sci. Rep.* 9.
- Wang, N., Zhang, K., Shen, X., Wang, Y., Li, J., Li, C., Mao, J., Malinka, A., Zhao, C., Russell, L.M., Guo, J., Gross, S., Liu, C., Yang, J., Chen, F., Wu, L., Chen, S., Ke, J., Xiao, D., Zhou, Y., Fang, J., Liu, D., 2022. Dual-field-of-view high-spectral-resolution lidar: simultaneous profiling of aerosol and water cloud to study aerosol-cloud interaction. *Proc. Natl. Acad. Sci. U. S. A.* 119 (10).
- Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., Cribb, M., 2019. Estimating 1-km-resolution PM<sub>2.5</sub> concentrations across China using the space-time random forest approach. *Remote Sens. Environ.* 231.
- Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T., Cribb, M., 2021a. Reconstructing 1-km-resolution high-quality PM<sub>2.5</sub> data records from 2000 to 2018 in China: spatiotemporal variations and policy implications. *Remote Sens. Environ.* 252, 112136.
- Wei, J., Li, Z., Pinker, R.T., Wang, J., Sun, L., Xue, W., Li, R., Cribb, M., 2021b. Himawari-8-derived diurnal variations in ground-level PM<sub>2.5</sub> pollution across China using the fast space-time light gradient boosting machine (LightGBM). *Atmos. Chem. Phys.* 21 (10), 7863–7880.
- Wu, J., Yao, F., Li, W., Si, M., 2016. VIIRS-based remote sensing estimation of ground-level PM<sub>2.5</sub> concentrations in Beijing-Tianjin-Hebei: a spatiotemporal statistical model. *Remote Sens. Environ.* 184, 316–328.
- Xiao, D., Wang, N., Shen, X., Landulfo, E., Zhong, T., Liu, D., 2020. Development of ZJU high-spectral-resolution Lidar for aerosol and cloud: extinction retrieval. *Remote Sens.* 12 (18).
- Xing, J., Mathur, R., Pleim, J., Hogrefe, C., Gan, C.M., Wong, D.C., Wei, C., Gilliam, R., Pouliot, G., 2015. Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe. *Atmos. Chem. Phys.* 15 (5), 2723–2747.
- Yan, X., Zang, Z., Luo, N., Jiang, Y., Li, Z., 2020. New interpretable deep learning model to monitor real-time PM<sub>2.5</sub> concentrations from satellite data. *Environ. Int.* 144.
- Zang, L., Mao, F., Guo, J., Wang, W., Pan, Z., Shen, H., Zhu, B., Wang, Z., 2019. Estimation of spatiotemporal PM<sub>1.0</sub> distributions in China by combining PM<sub>2.5</sub> observations with satellite aerosol optical depth. *Sci. Total Environ.* 658, 1256–1264.
- Zhang, T.H., Zhu, Z.M., Gong, W., Zhu, Z.R., Sun, K., Wang, L.C., Huang, Y.S., Mao, F.Y., Shen, H.F., Li, Z.W., Xu, K., 2018. Estimation of ultrahigh resolution PM<sub>2.5</sub> concentrations in urban areas using 160 m Gaofen-1 AOD retrievals. *Remote Sens. Environ.* 216, 91–104.
- Zhao, H.J., Che, H.Z., Wang, Y.Q., Dong, Y.S., Ma, Y.J., Li, X.X., Hong, Y., Yang, H.B., Liu, Y.C., Wang, Y.F., Gui, K., Sun, T.Z., Zheng, Y., Zhang, X.Y., 2018. Aerosol vertical distribution and typical air pollution episodes over northeastern China during 2016 analyzed by ground-based lidar. *Aerosol Air Qual. Res.* 18 (4), 918–937.