

Environmental Research Communications



LETTER

OPEN ACCESS

RECEIVED

30 September 2019

REVISED

4 December 2019

ACCEPTED FOR PUBLICATION

13 February 2020

PUBLISHED

24 February 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Identify the contribution of elevated industrial plume to ground air quality by optical and machine learning methods

Limin Feng^{1,2}, Ting Yang¹ , Dawei Wang¹, Zifa Wang¹, Yuepeng Pan¹, Ichiro Matsui³, Yong Chen¹, Jinyuan Xin¹ and Huili Huang¹

¹ State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, People's Republic of China

² University of Chinese Academy of Sciences, Beijing, 100049, People's Republic of China

³ National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, 305-8506, Japan

E-mail: tingyang@mail.iap.ac.cn

Keywords: elevated stack plume, PM_{2.5} concentration, particle scavenging, machine learning

Supplementary material for this article is available [online](#)

Abstract

Regional severe haze caused by atmospheric particle explosion is one of the biggest environmental problems in China that has yet to be fully understood. This research managed to find the linkage between diversified shapes of heavy industrial stack plume (HISP) and local ground particle concentration. We used two optical methods: LIDAR and auto-shoot camera, to catch the HISP's vertical shape, and two machine learning models: binary classification and decision tree, to find the quantitative relationship between the HISP's shape and PM_{2.5} concentration. The PM_{2.5} concentration correlated to the polygon length (PL) of HISP's shape with a logistic function. With a plume length more than twice the height of stack, the spread of HISP's shape accompanied with PM_{2.5} concentration decreasing to $< 100 \mu\text{g m}^{-3}$. The residence time of HISP's particles was longer (> 20 h) under uniform offshore dispersion than that in heterogeneous wind field, when the footprint of HISP was estimated to be > 7 km. We acquired a decision tree model to yield an exact prediction of PM_{2.5} concentration, in which the HISP's length played a statistically significant role. Though the plume shape is just one of the easy-to-use indicators of complex meteorological condition, it is still practical for policy makers to identify the particle pollution caused by the elevated sources in the fastest way.

1. Introduction

With uniform state planning, the Chinese government has produced a spectacular improvement of air quality in the past five years [1–3]. However, regional severe haze events in which fine particle (diameter $< 2.5 \mu\text{m}$) concentration reached to $> 200 \mu\text{g m}^{-3}$ still plagued north China, especially in wintertime [4, 5]. A better understanding of the regional air pollution is urgently needed for developing follow-through remediation. Notably, with the intensive energy consumption and air pollutant emission, China's heavy industries are producing 2–3 billion tons of iron/steel and consuming 3–4 billion tons of coke/coal per year, accounting for half of the total world output [6]. These heavy industrial plants, where stack plumes are usually treated with low efficiency waste gas cleaning, are centrally distributed in north China [5, 7]. From a regional view, the heavy industrial stack plume (HISP) could rise to a height of more than 1000 m, and inclines to transport from the elevated emission source to the downwind area [8, 9], interacts with thermals from the land, and dilutes rapidly to the ground [10, 11]. From a local view, the HISP mixes with ambient air masses just several hours after it is emitted [12, 13]. The visible white or gray shape of HISP appears because of water vapor condensation but the actual amount of dispersing particles is far beyond its envelope. Therefore, HISP induces a major human health risk to local and downwind areas, and it is important for policy makers to quickly determine how much the role of HISP plays in air pollution.

Previously, numerical model was the most commonly employed method to evaluate the HISP impacts. In classical physical model, such as Gaussian plume model using prescribed meteorological data, the plume's particles disperse under the assumption of a uniform state with homogeneous wind field. However, these assumptions are deficient when the meteorology is spatially and temporally heterogeneous prevailing in the most common conditions [14]. Assuming instantaneous vertical mixing of plume particles, Lagrangian model also produces inaccurate results (typically underestimated) when entrainment rate is large or vertical spread speed is small [15, 16]. Eulerian numerical models perform even worse than the Lagrangian model, since these models cannot resolve the sub-grid processes [16, 17]. Besides, classical physical approaches are computationally expensive, dependent to large databases of input parameters which may not be complete [18]. There is a growing demand for decision-makers to provide rapid, economical and effective solutions for air pollution control [19]. The simplest way to track changeable HISP is using optical method such as cameras catching the boundary of elevated plume [20]. The vertical variability of elevated plume has been less well studied than that in the horizontal plane [21]. Optical photo methods enable sensing over a wide span of space to catch HISP's dispersion and transport [22]. When the bottom envelope of HISP is close to the ground, chances are that the plume's particles increase ground $PM_{2.5}$ concentration since they mainly distribute in fine size [23]. On top of optical methods, more practitioners resort to data-driven methods, such as machine learning, which can predict air pollution and identify the key factors without consider complex physical and chemical processes [24–26]. Note that optical and data-driven methods are generally case-specific, produce limited repeatability, which is the irreplaceable advantage of classical physical models [25].

The dispersion of HISP is determined by meteorological conditions and the stack parameters, such as stack height, stack diameter, plume exit temperature, and plume exit velocity. The plume shape is one of the easy-to-use indicators of complex meteorological condition. We choose a coastal industrial area to apply the optical and machine learning methods using the indicator of plume shape. Coastal regions are often prefer sites for heavy industry development and have high emission loads, extending over hundreds of kilometers [27]. In the recent years of remediation policies, heavy industry such as *Iron and Steel Foundry* (ISF) in the north China plain is moving to coastal areas to avoid fumigation to this largest population center. Specifically, industrial smoke emission in Shandong province ranks the highest (1.2 million tons) among all provinces in China [28]. Rizhao is a coastal city in Shandong province, with numerous ISF plants and a still increasing productive capacity because of Chinese uniform planning. In this paper, we enter the fray of air pollution control by catching the HISP's shape, which is emitted from ISF located in the industrial area of Rizhao city. We managed to find the relationship between HISP's shape, meteorological conditions, and ground $PM_{2.5}$ concentration. Doing so can be helpful for policy-makers to identify the air pollution caused by industrial sources with real-time evaluation.

2. Materials and methods

In September 2018, we launched a campaign at an urban site in an ISF in Rizhao (35.124 °N, 119.311 °E). The ISF's stacks are located in the dense production area and the number of stacks is about 200. Satellite maps of ISF's heavy stack plumes and their affected area can be found in this manuscript (figure S1 is available online at stacks.iop.org/ERC/2/021005/mmedia). During the observation in early autumn, the coal burning for residential heating can be ignored. At the ISF site, a dual-wavelength (NIES-MIE 1064, 532 nm) depolarization LIDAR was set up for discerning the industrial plumes. LIDAR was developed by the National Institute for Environmental Studies (Japan). More details can be found therein [29, 30]. In March 2019, we set up an auto-shoot camera fixed at the ISF site to catch the different types of HISP's shapes. After obtaining the photos, we used the software Labelme (GitHub: <https://github.com/wkentaro/labelme>) to denote the HISP's shape and then we calculated the polygon length of the denoted shape. Some representative photos can be seen in figure 1. The HISP's shape was denoted by hand, thus there were operation errors during the labeling process. The $PM_{2.5}$ concentration data comes from a local monitoring station in the ISF. The meteorological parameters at different heights are available on the public website of European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data (ERA-interim, 0.125° × 0.125°). The international weather station (ID = 549450) for assimilating the ECMWF model is about 20 km from the ISF site. We abstracted the profile of Air temperature (T) and wind (U, V) in the ISF site. The related denoted files (.json) and other types of data are available in the GitHub website created by Limin Feng for public review (https://github.com/Limin-Feng1993/Data_Plume_Shape_and_Particle_Concentration).

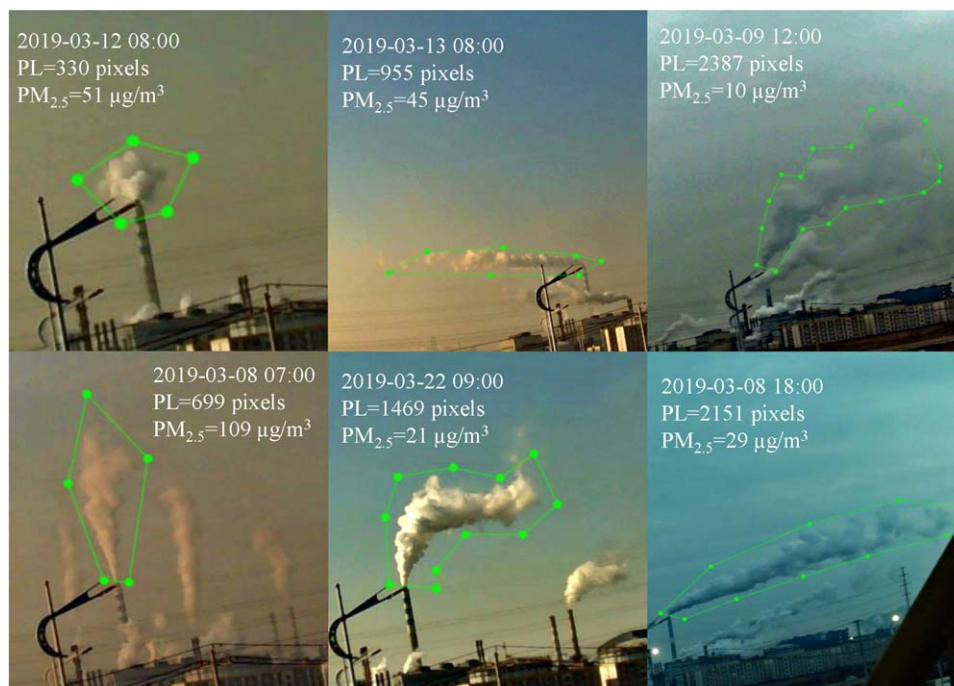


Figure 1. Different polygon length (pixels) of the stack plumes denoted by hand and PM_{2.5} concentration ($\mu\text{g m}^{-3}$).

3. Results and discussion

3.1. HISP's shape caught by extinction signals

The scattering components of HISP's particles, such as sulfate and nitrite, induce a large extinction coefficient in the upper air and is responsible for the degradation of visibility [31]. In September 2018, we caught 4 episodes with significant extinction signals at >200 m height, indicating that different types of HISP passed across the ISF site.

3.1.1. Episode-1: HISP's trapping by stable air

Plume trapping occurs when a plume is capped by a deep lid of stable air, which is frequent on cloudy days [32]. Under low cloud cover on September 2, we caught significant extinction signals at 500–1000 m height, named episode-1 (figure 2(a)). The cloud signal is marked as blank areas in the figure. The cloud base height was about 1000 m height, and the upward dispersion of HISP was suppressed. The extinction coefficient of HISP's particles was close to 1. Due to the interaction of low clouds and stack plumes, the internal mixing of HISP's particles was enhanced, inducing significant radiation force and extinction [33]. The static air appeared at <1000 m height with a wind speed close to zero, while the ground PM_{2.5} concentration decreased from $22 \mu\text{g m}^{-3}$ to $7 \mu\text{g m}^{-3}$. The airborne particle concentration didn't accumulate, suggesting that the surface air was lack of external pollutant input, neither from ground nor HISP. Static condition in the surface air warded off the upper sinking HISP.

3.1.2. Episode-2 and episode-3: HISP's spreading in homogeneous wind field

On September 10, we caught the HISP passing by the observation site with significant signals of aerosols extinction at 500–1500 m height, named as episode-2 (figure 2(b)). The bottom boundary of the HISP was lower than 200 m height (close to the stack height), indicating the HISP's sinking and fumigation. In episode-2, the west offshore wind dominated 0–3000 m height with a wind speed $> 2 \text{ m s}^{-1}$, inducing strong disturbance to the plume aerosols. Accordingly, the ground PM_{2.5} concentration ranged from 24 – $71 \mu\text{g m}^{-3}$ during episode-2, suggesting that ground aerosols mixed with HISP particles. Comparing episode-1 and episode-2, we conclude that the unstable air at < 1000 m height intensified the HISP's downward dispersion, while the stable air at >1000 m height suppressed the HISP's upward dispersion.

Episode-3 occurred from September 7 to September 8, when we caught two HISPs passing by the upper air one after the other. It can be seen in figure 2(b) that the HISPs spread within 500–3000 m height. Both HISPs can be continuously observed over 20 h. When episode-3 occurred, the north wind dominated 0–3000 m height lasting for more than 48 h. The homogeneous northerly winds created a distinct clean layer between HISP's

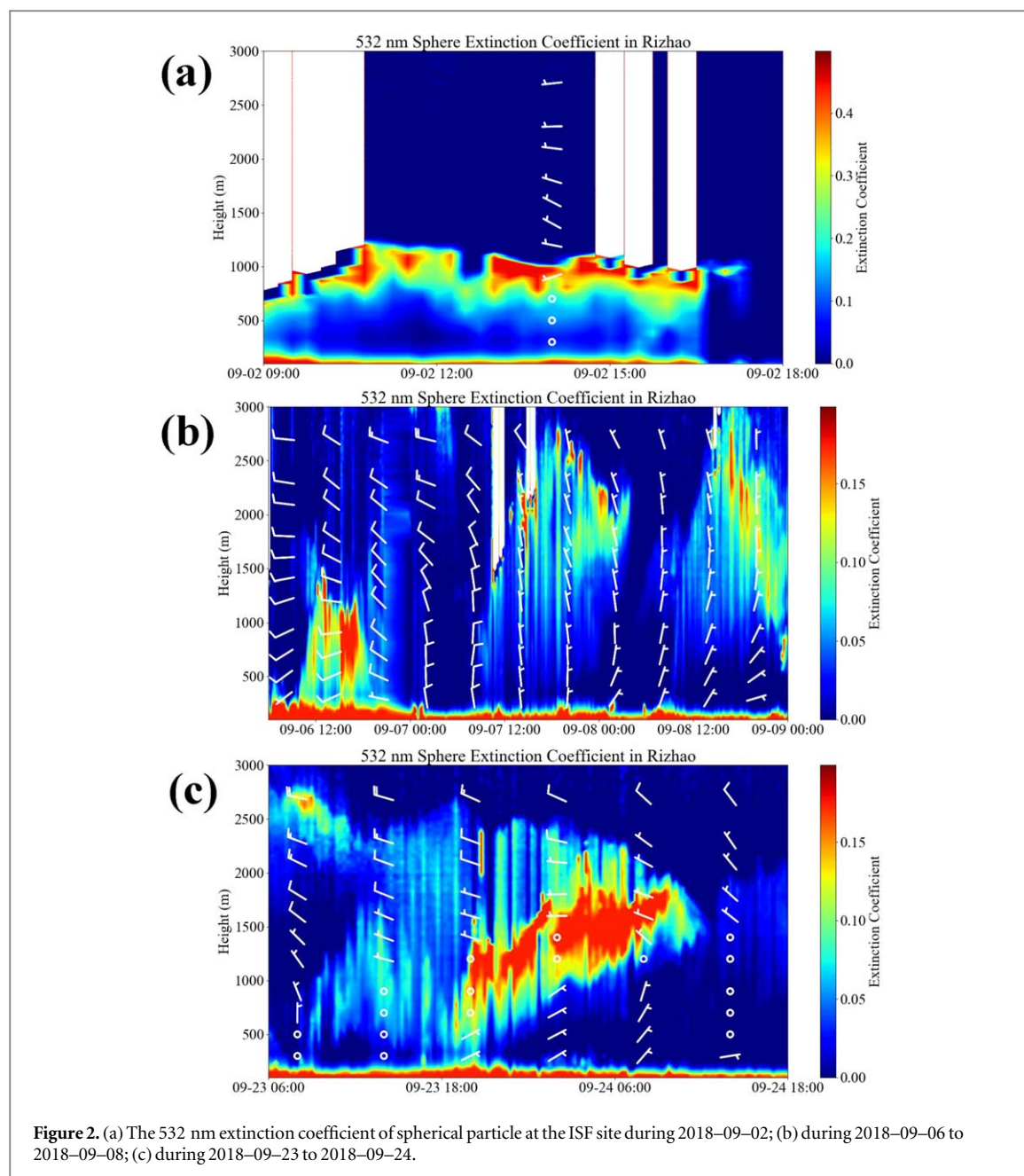


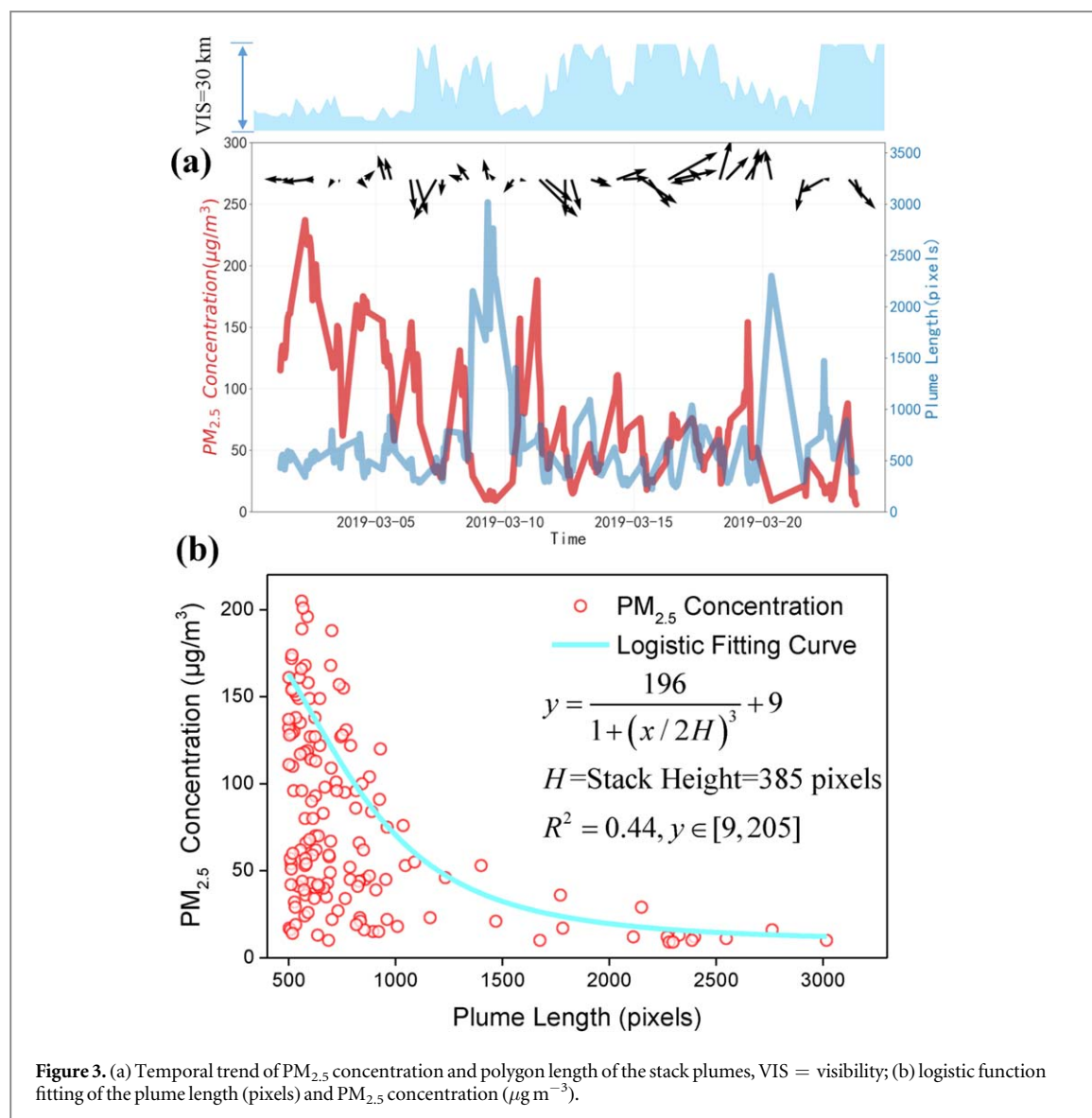
Figure 2. (a) The 532 nm extinction coefficient of spherical particle at the ISF site during 2018–09–02; (b) during 2018–09–06 to 2018–09–08; (c) during 2018–09–23 to 2018–09–24.

particles and ground aerosols, accordingly, ground $\text{PM}_{2.5}$ concentration ranged from $7\text{--}12\ \mu\text{g m}^{-3}$ during episode-3. In just 6 h, the two HISP's spread over a span of 2000 m, thus the dispersion rate of HISP was estimated as $\sim 6\ \text{m min}^{-1}$ in the vertical direction. Given that the plumes tilt at some angle rather than flat or straight, we assume that the transport speed of HISP in the horizontal direction was comparable to the vertical direction (typically larger). Thus the footprint of the passing-by HISP was estimated to be at least 7 km (residence time as 20 h).

3.1.3. Episode-4: HISP's spreading in heterogeneous wind field

Under sea-land breeze (typically $< 1000\ \text{m}$ height), temporally unstable and spatially heterogeneous wind causes a vertical shear of the plume shape [34, 35]. Episode-4 occurred on September 23, during which the wind direction changed from east in $0\text{--}1000\ \text{m}$ to west in $1000\text{--}3000\ \text{m}$ after 18:00 (figure 2(c)). During daytime, the plume spread over the $500\text{--}3000\ \text{m}$ height under the robust west land breeze in $1000\text{--}3000\ \text{m}$ and static air at $0\text{--}1000\ \text{m}$ height. However, at 18:00, east sea breeze began to dominate the air at $< 1000\ \text{m}$ height, and the bottom boundary of the HISP was sinking to $< 300\ \text{m}$. Then during nighttime, the HISP's shape was shrinking, distinguished from ground aerosols, and disappeared at $\sim 1500\ \text{m}$ height in the next morning.

After 18:00 on September 23, the extinction coefficient in the HISP's center line increased from ~ 0.08 to > 0.15 . The extinction is mainly determined by coarse particles with a diameter $> 1\ \mu\text{m}$ [36, 37]. The 532 nm



extinction coefficient showed in figure 2 is mainly determined by the upper particles with a diameter $> 532 \text{ nm}$, and increase as particles grow larger. Therefore, we infer that plume's particles shifted to a larger size and were then continuously scavenged after 18:00 on September 23. The enlarged particle size can be explained by hygroscopic growth, since coagulation rate for super-micron particles is too low to affect particle growth [38]. Substantial particle volume growth was reported to associate with sulfate-rich plumes [39]. The particle growth in HISP was enhanced after the condensation of SO_2 and taking up droplets. As a result, the in-cloud scavenging amplified the absorption and scattering of the HISP meanwhile reduced the residence time of plume's particles.

3.2. HISP's shape caught by visible photos

To further determine the impact of HISP's shape to ground particle concentration, we denoted and calculated the polygon length of the visible plume from campaign in March 2019, and attempted to discover the quantitative relationship between them. The polygon length means the maximum distance of two points in all vertexes (basically the diagonal line). The long/broad HISP contains large amount water droplets and liquid particles with long-lasting dispersion.

3.2.1. The non-linear correlation between HISP's shape and $PM_{2.5}$ concentration

From the temporal trend of plume length and $PM_{2.5}$ concentration, there seems to be a negative correlation between the two parameters, especially in the days with long HISP, (figure 3(a)). However, this negative correlation is non-linear and cannot be regarded as a causal relationship. We can see in figure 3(a) that $PM_{2.5}$ concentration was also controlled by synoptic conditions inducing different visibilities and wind fields. For example, during March 1 to March 5 and March 11 to March 12, the visibility was less than 5 km caused by the sea fog and low clouds (see typical photos in figure S3). Accordingly, the $PM_{2.5}$ concentration stayed at a high

level of $>100 \mu\text{g m}^{-3}$. The cloud/fog processes of HISP's particles and ground aerosols were enhanced during these low visibility weathers with intensified phase transformation and hygroscopic growth [27, 40].

At first we simplified the question by using the logistic function to fit the polygon length (PL) of HISP's shape and $\text{PM}_{2.5}$ concentration. Logistic function is a generalized Boltzmann distribution, also known as Gibbs distribution, which is the probability distribution of particles in a nearly isolated system under heat balance. Here we get: $\frac{196}{1 + (x/2H)^3} + 9$, $y \in [9, 205]$ H represents the stack height which was about 200 m and 385 pixels in the photo (figure 3(b)). The value of $9 \mu\text{g m}^{-3}$ represents the background value of $\text{PM}_{2.5}$ concentration under the most favorable condition, and $205 \mu\text{g m}^{-3}$ represents the maximum impact of HISP. Note that the distribution of PL versus $\text{PM}_{2.5}$ concentration is only half of the logistic function. This is because the HISP's length cannot be as small as zero and always has a certain length. During March 1 to March 5, sea fog appears at the ISF site, and the plume length was <500 pixels in the photos with a large operation error. Therefore, we only choose the HISP's shape with a polygon length of >500 pixels as credible results for fitting.

In machine learning, the logistic function cannot be regarded as a regression line, instead, it is a binary classification method [41]. The logistic function classifies the distribution into two groups. In the one group with plume length $>2H$, the spreading of HISP's shape accompanied with the $\text{PM}_{2.5}$ concentration decreasing to $<100 \mu\text{g m}^{-3}$. In the other group with plume length $<2H$, the $\text{PM}_{2.5}$ concentration ranged from $10 \mu\text{g m}^{-3}$ to $200 \mu\text{g m}^{-3}$. In other words, the HISP's shape has no significant correlation with $\text{PM}_{2.5}$ concentration, which is the main source of fitting errors. As formerly said, there are other factors dominating ground $\text{PM}_{2.5}$ concentration in addition to the HISP's shape. Both the HISP's particles and ground particles can be regarded as subset samples of mixing layer aerosols. Therefore, the collaborative change of these two subset aerosols only occurred under robust mixing. While under the weak diffusion conditions symbolized by short HISP, the ground aerosols may originate from road dust from the transport of raw materials for ISF plants and residential combustion in winter. Thus the controlling factors of $\text{PM}_{2.5}$ concentration in the second group were complex.

3.2.2. The multiple factors controlling ground $\text{PM}_{2.5}$ concentration

Since the logistic function still lags far behind our requirement for predicting $\text{PM}_{2.5}$ concentration, we take wind and temperature profiles into consideration, and then adopt decision tree regression (another popular machine learning method) to make the prediction transparent and precise [42]. We reduce the temporal resolution of data to per 6 h to match the ECMWF meteorological data (U, V, and T at 850 hPa, 950 hPa, and 1000 hPa, respectively). The simulation value and observation values can be seen in figure 4(a). The root mean square error (RMSE) of the decision tree regression is $10 \mu\text{g m}^{-3}$ for the training set and $19 \mu\text{g m}^{-3}$ for the test set. The large RMSE in the test set is mainly caused by several far over-estimated values (blue outliers in figure 4(a)). In general, the decision tree model performs well in predicting the $\text{PM}_{2.5}$ concentration with acceptable error and R-square value (>0.8).

The depth of decision tree is 6 (choose the branch number with the highest prediction precision), which is tedious to explain (figure 4(b)). Subscripts of U and T represent the different pressure levels (hPa), and the negative value of U represents westward wind (onshore). However, we only need to focus two main trunks of the decision tree: the left one controlled by strong eastward wind (U) and short HISP, the right one controlled by long HISP. The left one (with red border) is: $U_{850} < -5.4 \text{ m s}^{-1} \rightarrow \text{PL} < 509 \text{ pixels} \rightarrow \text{PM}_{2.5} > 176 \mu\text{g m}^{-3}$. The right one (with blue border) is: $U_{850} > -5.4 \text{ m s}^{-1} \rightarrow \text{PL} > 796 \text{ pixels} (\sim 2H) \rightarrow \text{PM}_{2.5} < 67 \mu\text{g m}^{-3}$. The former trunk tells us that the robust onshore wind and short HISP were related to severe particle pollution at ground level. The onshore east wind brings back the dispersed HISP, causes secondary plume fumigation, and produces advection fog to the coastal area.

The latter trunk tells us that under the offshore wind and long/broad HISP, the $\text{PM}_{2.5}$ concentration was at a low level and harmless to humans. In this case, HISP's particles were scavenged faster under the integrated effect of offshore dispersion and in-plume droplet's cleaning. It is ill-formed to keep faith with ground-level $\text{PM}_{2.5}$ concentration leading to take redundant effort to control emission sources. This decision tree model yields an exact prediction of $\text{PM}_{2.5}$ concentration using the HISP's shape and wind profile, which is practical for policy makers to identify the particle pollution caused by elevated sources or meteorological conditions in the fastest way.

4. Conclusions

In this research, we performed an analysis of the contribution of heavy industrial stack plume to local ground particle concentration. An auto-shoot camera and LIDAR were set up to catch the passing-by HISP. Here we suggest that:

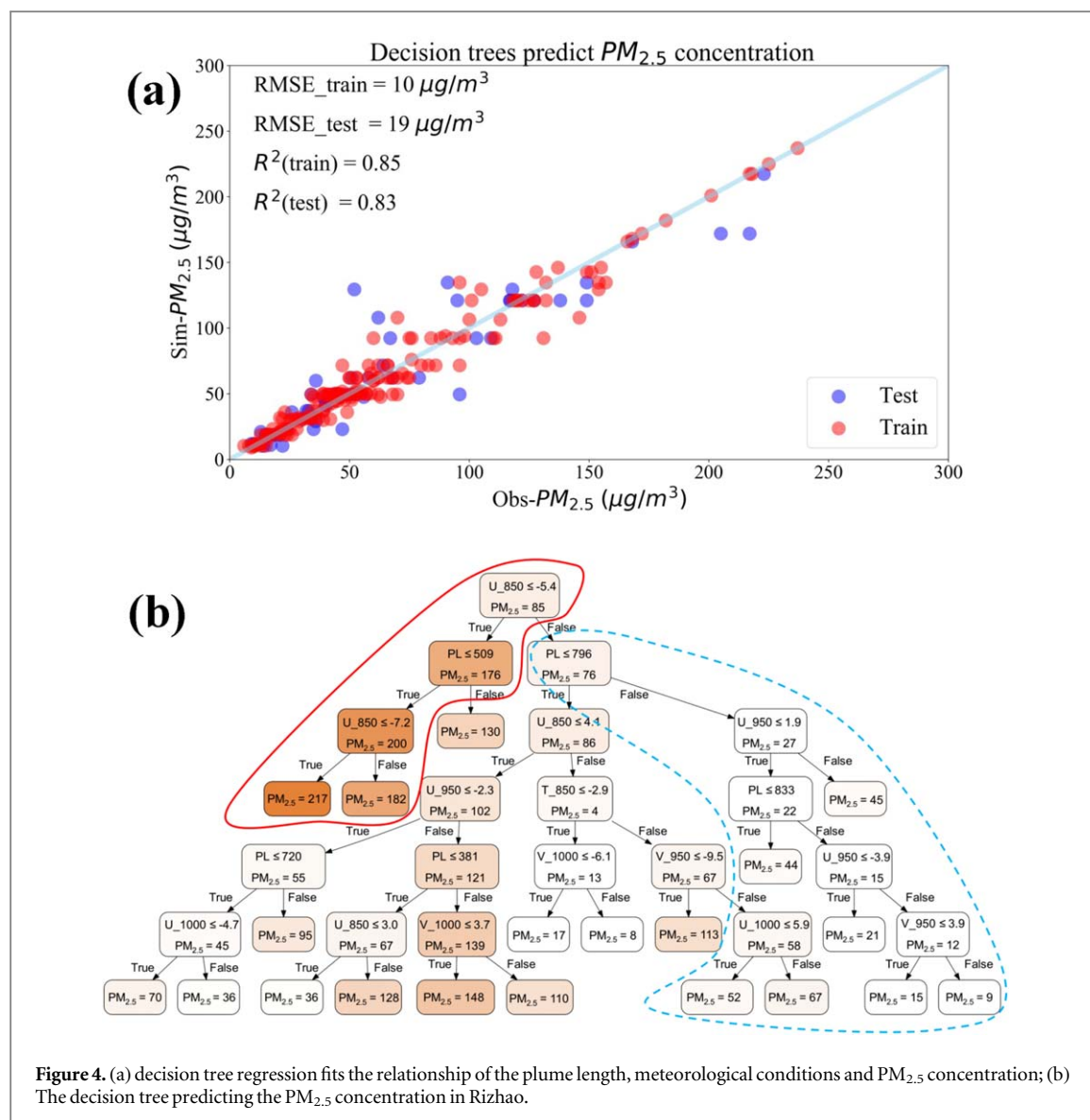


Figure 4. (a) decision tree regression fits the relationship of the plume length, meteorological conditions and $PM_{2.5}$ concentration; (b) The decision tree predicting the $PM_{2.5}$ concentration in Rizhao.

- The ground $PM_{2.5}$ concentration correlated to the polygon length of HISP with a logistic function. With a plume length more than twice the height of stack, the spreading of HISP's shape accompanied with the substantially decreased $PM_{2.5}$ concentration. The longer elevated stack plumes induce a smaller detrimental impact on ground air quality.
- The residence time of HISP's particles was longer (>20 h) under uniform offshore dispersion than that in heterogeneous wind field, when the footprint of HISP was estimated to be >7 km. Whether the vertical wind field is homogeneous or not determines the plume's dispersing boundary.
- We acquired a decision tree model to yield an exact prediction of $PM_{2.5}$ concentration, which confirms that the HISP's shape played a fundamental role in determining the local atmospheric particle concentration. It is an economic way to evaluate industrial air pollution using the stack plume's shape.

Acknowledgments

We would like to thank the support from Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA19040203). National Key Research and Development Project (No. 2017YFC0209903)

Data availability

The data are available in the GitHub website created by Limin Feng for public review (https://github.com/Limin-Feng1993/Data_Plume_Shape_and_Particle_Concentration).

Author contributions

LF, TY and ZW designed the whole structure of this work. YP, TY, DW, IM and HL prepared the ground monitoring data and LIDAR data, respectively. JX and YC prepared the wind profile data. LF analyzed the data. LF and TY prepared the manuscript with contributions from all-authors.

Competing financial interests

The authors declare no competing financial interests.

ORCID iDs

Ting Yang  <https://orcid.org/0000-0001-5605-0654>

References

- [1] Zhao S *et al* 2018 Temporal dynamics of SO₂ and NO_x pollution and contributions of driving forces in urban areas in China *Environ. Pollut.* **242** 239–48
- [2] An Z *et al* 2019 Severe haze in northern China: a synergy of anthropogenic emissions and atmospheric processes *Proc. of the National Academy of Sciences* **116**, 8657–66
- [3] Wang S and Hao J 2012 Air quality management in China: issues, challenges, and options *J. Environ. Sci.* **24** 2–13
- [4] Cai W, Li K, Liao H, Wang H and Wu L 2017 Weather conditions conducive to Beijing severe haze more frequent under climate change *Nat. Clim. Change* **7** 257–62
- [5] Xu H *et al* 2019 Spatial and temporal distribution, chemical characteristics, and sources of ambient particulate matter in the Beijing-Tianjin-Hebei region *Sci. Total Environ.* **658** 280–93
- [6] National Bureau of Statistics of China 2019 China statistical yearbook 2018 China Statistics Press ISBN 978-7-5037-8587-0
- [7] Li R *et al* 2019 Air pollution characteristics in China during 2015–2016: spatiotemporal variations and key meteorological factors *Sci. Total Environ.* **648** 902–15
- [8] Zaveri R A *et al* 2010 Overnight atmospheric transport and chemical processing of photochemically aged Houston urban and petrochemical industrial plume *Journal of Geophysical Research: Atmospheres* **115**
- [9] Leoni C, Hovorka J, Dočekalová V, Cajthaml T and Marvanová S 2016 Source impact determination using airborne and ground measurements of industrial plumes *Environmental Science & Technology* **50** 9881–8
- [10] Tsuang B-J *et al* 2003 Quantification on the source/receptor relationship of primary pollutants and secondary aerosols by a Gaussian plume trajectory model: Part II. Case study *Atmos. Environ.* **37** 3993–4006
- [11] Meroney R, Cermak J and Yang B 1975 Modeling of atmospheric transport and fumigation at shoreline sites *Boundary Layer Meteorol.* **9** 69–90
- [12] Marris H *et al* 2012 Fast changes in chemical composition and size distribution of fine particles during the near-field transport of industrial plumes *Sci. Total Environ.* **427–428** 126–38
- [13] Mylläri F *et al* 2016 New particle formation in the fresh flue-gas plume from a coal-fired power plant: effect of flue-gas cleaning *Atmos. Chem. Phys.* **16** 7485–96
- [14] Segal M, Mahrer Y and Pielke R A 1982 A numerical model study of plume fumigation during nocturnal inversion break-up *Atmospheric Environment* (1967) **16** 513–9
- [15] Lunar A K and Sawford B L 1996 An examination of existing shoreline fumigation models and formulation of an improved model *Atmos. Environ.* **30** 609–20
- [16] Kim Y, Seigneur C and Duclaux O 2014 Development of a plume-in-grid model for industrial point and volume sources: application to power plant and refinery sources in the Paris region *Geoscientific Model Development* **7** 569–85
- [17] Tsuang B-J 2003 Quantification on the source/receptor relationship of primary pollutants and secondary aerosols by a Gaussian plume trajectory model: Part I—Theory *Atmos. Environ.* **37** 3981–91
- [18] Jiang P, Dong Q and Li P 2017 A novel hybrid strategy for PM_{2.5} concentration analysis and prediction *J. Environ. Manage.* **196** 443–57
- [19] Moustiris K P, Ziomas I C and Paliatatos A G 2010 3-Day-ahead forecasting of regional pollution index for the pollutants NO₂, CO, SO₂, and O₃ using artificial neural networks in Athens, Greece *Water Air Soil Pollut.* **209** 29–43
- [20] Weibring P, Andersson M, Edner H and Svanberg S 1998 Remote monitoring of industrial emissions by combination of lidar and plume velocity measurements *Applied Physics B: Lasers Optics* **66** 383–8
- [21] Huang L, Jiang J H, Tackett J L, Su H and Fu R 2013 Seasonal and diurnal variations of aerosol extinction profile and type distribution from CALIPSO 5-year observations *Journal of Geophysical Research: Atmospheres* **118** 4572–96
- [22] Gross K C, Bradley K C and Perram G P 2010 Remote identification and quantification of industrial smokestack effluents via imaging fourier-transform spectroscopy *Environmental Science & Technology* **44** 9390–7
- [23] Choël M, Deboudt K and Flament P 2010 Development of time-resolved description of aerosol properties at the particle scale during an episode of industrial pollution plume *Water, Air, Soil Pollution* **209** 93–107
- [24] Chen G *et al* 2018 A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information *Sci. Total Environ.* **636** 52–60
- [25] Cabaneros S M S, Calautit J K and Hughes B R 2019 A review of artificial neural network models for ambient air pollution prediction *Environ. Modell. Softw.* **119** 285–304
- [26] Reichstein M *et al* 2019 Deep learning and process understanding for data-driven Earth system science *Nature* **566** 195–204
- [27] Kourtidis K, Stathopoulos S, Georgoulas A, Alexandri G and Rapsomanikis S 2015 A study of the impact of synoptic weather conditions and water vapor on aerosol–cloud relationships over major urban clusters of China *Atmos. Chem. Phys.* **15** 10955–64
- [28] National Bureau of Statistics of China 2019 China statistical yearbook on Environment 2018 China Statistics Press ISBN 978-7-5037-8780-5

- [29] Sugimoto N *et al* 2002 Observation of dust and anthropogenic aerosol plumes in the Northwest Pacific with a two-wavelength polarization lidar on board the research vessel Mirai *Geophys. Res. Lett.* **29** 7-1-7-4
- [30] Yang T *et al* 2017 Technical note: boundary layer height determination from lidar for improving air pollution episode modeling: development of new algorithm and evaluation *Atmos. Chem. Phys.* **17** 6215-25
- [31] Dubovik O *et al* 2002 Variability of absorption and optical properties of key aerosol types observed in worldwide locations *J. Atmos. Sci.* **59** 590-608
- [32] Lyons W A and Cole H S 1973 Fumigation and plume trapping on the shores of lake michigan during stable onshore flow *J. Appl. Meteorol.* **12** 494-510
- [33] Penner J, Chuang C and Grant K 1998 Climate forcing by carbonaceous and sulfate aerosols *Clim. Dyn.* **14** 839-51
- [34] Luhar A K and Young S A 2002 Dispersion moments of fumigating plumes—lidar estimates and PDF model simulations *Boundary Layer Meteorol.* **104** 411-44
- [35] Luhar A K 2002 The influence of vertical wind direction shear on dispersion in the convective boundary layer, and its incorporation in coastal fumigation models *Boundary Layer Meteorol.* **102** 1-38
- [36] Elias T *et al* 2009 Particulate contribution to extinction of visible radiation: pollution, haze, and fog *Atmos. Res.* **92** 443-54
- [37] Li Z *et al* 2013 Aerosol physical and chemical properties retrieved from ground-based remote sensing measurements during heavy haze days in Beijing winter *Atmos. Chem. Phys.* **13** 10171-83
- [38] Badger C L *et al* 2006 Phase transitions and hygroscopic growth of aerosol particles containing humic acid and mixtures of humic acid and ammonium sulphate *Atmos. Chem. Phys.* **6** 755-68
- [39] Brock C A *et al* 2003 Particle growth in urban and industrial plumes in Texas *J. Geophys. Res.* **108**
- [40] Andreae M O and Rosenfeld D 2008 Aerosol-cloud-precipitation interactions. Part 1. The nature and sources of cloud-active aerosols *Earth Sci. Rev.* **89** 13-41
- [41] King G and Zeng L 2001 Logistic regression in rare events data *Political Analysis* **9** 137-63
- [42] Tso G K F and Yau K K W 2007 Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks *Energy* **32** 1761-8