General comments

This work concerns the development of a deep learning model to estimate PM2.5 concentrations from satellite data also under cloudy conditions. Although the topic addressed is interesting and quite original, the manuscript is not in a good shape for being published. I share most of the comments from Anonymous Referee #1 in the lack of clarity of some sections, especially as for the methods and the many missing details. In the Introduction section, lack of knowledge of the main physicochemical processes affecting PM production and removal seem to emerge. This is not negligible, since these can be connected with the discussion of the results obtained. Indeed, also the discussion section right now seems more like a sort of celebration of the results achieved, while a clear discussion on the results obtained, including reasons (processes?) and references, is totally missing. Clear linkages between the different variables that are obtained from ground based stations (PM concentrations, in ug m^{-3}) and from satellites (AOT/AOD data) are also completely missing. Again, this is very useful for understanding the differences between the use and meaning of the variables, and why it is in any case not trivial to use satellite data for estimating PM concentrations at ground.

Reply:

- In the revised manuscript, we added the following background of $PM_{2.5}$, including how it is created, removed from the atmosphere and where it lingers in the atmosphere.
- "PM2.5 in the atmosphere is either emitted directly or formed from gaseous precursors through complex gas phase, aqueous phase or heterogeneous chemical reactions (Cheng et al., 2016; Gao et al., 2016). $PM_{2.5}$ consists of complex composition, including sulfate, nitrate, organic carbon, elemental carbon, soil dust, and sea salt (Gao et al., 2016). It can stay in the boundary layer for a few days and in the free troposphere for a few weeks (similar to ozone). It can be efficiently removed out of atmosphere by precipitation, which is the major atmospheric sink (Jacob and Winner, 2009)."
- "Aerosol optical depth (AOD) is the vertical integration of aerosol extincition from the surface to the top of the atmosphere, while ground-based instrument only measures $PM_{2.5}$ concentrations near the ground. When long-range transport of particles occurs above the ground, high AOD does not always coincide with high $PM_{2.5}$ concentrations (Hu et al., 2022). Besides, an opposite seasonality of AOD and surface $PM_{2.5}$ over China was identified due to aerosol hygroscopic growth (Xu et al., 2019)."

Jun Xu, Feng Han, Mingzhu Li, Zhongzhi Zhang, Du Xiaohui, Peng Wei, On the opposite seasonality of MODIS AOD and surface $PM_{2.5}$ over the Northern China plain, Atmospheric Environment, Volume 215, 2019, 116909, ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2019.116909.

Cheng, Y., Zheng, G., Wei, C., Mu, Q., Zheng, B., Wang, Z., Gao, M., Zhang, Q., He, K., Carmichael, G. and Pöschl, U., 2016. Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China. *Science advances*, *2*(12), p.e1601530.

Gao, M., Carmichael, G.R., Wang, Y., Saide, P.E., Yu, M., Xin, J., Liu, Z. and Wang, Z., 2016. Modeling study of the 2010 regional haze event in the North China

Plain. *Atmospheric Chemistry and Physics*, *16*(3), pp.1673-1691.
Jacob, D.J. and Winner, D.A., 2009. Effect of climate change on air quality. *Atmospheric environment*, *43*(1), pp.51-63.
Hu, Q., Liu, C., Li, Q., Liu, T., Ji, X., Zhu, Y., Xing, C., Liu, H., Tan, W. and Gao, M., 2022. Vertical profiles of the transport fluxes of aerosol and its precursors between Beijing and its southwest cities. *Environmental Pollution*, *312*, p.119988.

- We also rewrote the materials and methods, and elaborated on the rationality of input data and selection. We also discussed in detail the impact of each variable on the results and the limitations of the model.
- We evaluated the relationship between input variables and ground $PM_{2.5}$ through statistical analysis. And we use sensitivity analysis method to evaluate the impact of each input variable on the results in the model.
- According to the follow-up suggestions, we modified the discussion section and added conclusions. We can obtain the full coverage ground $PM_{2.5}$ concentration by mining the spatiotemporal relationship of multi-source heterogeneous data. Therefore, in cloud and night scenarios, we can speculate on ground $PM_{2.5}$ through spatiotemporal correlated satellite AOD and other data.
- "2 Materials and Methods
- We built a deep learning model to improve estimates of ground-level PM2.5 concentrations, particularly for regions without sampling sites, and for conditions (cloudy, hazy, nighttime, etc.) where satellite retrievals are not available. Our research area is concentrated in the Middle East of China (Figure. S3), with a time span of four years from 2017 to 2020. In Section 2.1, we introduced the input data of the model, in Section 2.2 and 2.3, we introduced the model structure and training verification methods, and in Section 2.4, we introduced a sensitivity analysis method used to quantitatively analyze the impact of the input data on the results. This method is used to open the black box model of neural network……".

Specific comments
Line 30: It is not clear which attempts you are referring to. Please rephrase.
Reply:

- We revise it in the manuscript.
- "However, satellite remote sensing AOD are hampered under cloudy/hazy conditions or during nighttime."

Line 31: "constrained" does not seem the most appropriate term in this context. Rephrase.
Reply:

- We revise it in the manuscript.
- "surface $PM_{2.5}$ cannot be obtained with satellite remote sensing under cloudy/hazy conditions or during nighttime."

Line 32: Add "In this work," before "We introduce".
Reply:

- We revise it in the manuscript.
- “In this work, we introduce a deep spatiotemporal neural”

Lines 33-34: Absolutely unclear what you mean by “We use sensitivity analysis and visualization technology to open the neural black box data model”: rephrase.

Reply:

- We revise it in the manuscript.
- “We quantified the quantitative impact of input variables on the results using sensitivity and visual analysis of the model.”

Lines 35-37: Details on the accuracy and errors of the method need also to be given here.

Reply:

- We revise it in the manuscript.
- “This technique provides ground-level $PM_{2.5}$ concentrations with high spatial resolution (0.01°) and 24-hour temporal coverage. In central and eastern China, the cross validation result R^2 is greater than 0.8 and RMSE is less than 26 (μg/m³). Better constrained spatiotemporal distributions of $PM_{2.5}$ concentrations will help improve health effects studies, atmospheric emission estimates, and predictions of air quality.”

Line 40: What do you mean by “impediments to human health”? Revise.

Reply:

We revise it in the manuscript.

- “Ambient particles raise worldwide concerns due to adverse effects on human health”

Lines 41-42: Please explain better, quite obscure the meaning of this sentence.

Reply:

- We revise it in the manuscript.
- “aerosols affect weather and climate system by scattering and absorbing radiation from the ground and the sun. Besides, they serve as cloud condensation nuclei (CCN) or ice nuclei (IN) to change size of cloud droplets and cloud optical properties.”

Lines 44-45: Not totally true, considering that PM2.5 (and PM10) precursors are in some cases not regulated (consider for instance ammonium deriving from ammonia) or limitations are due to the fact that precursors themselves are pollutants/toxic. Also, secondary aerosols are transported from long-range. So, this sentence contains many technical faults which need to be addressed and better clarified as the formation of secondary aerosols is key and one of the main reasons we have difficulties in addressing limit PM concentration values worldwide.

Reply:

- We revise it in the manuscript.
- “Accordingly, sources of $PM_{2.5}$ and $PM_{2.5}$ precursors are highly regulated in some

industrialized countries. As we gradually deepen our understanding of $PM_{2.5}$, more and more prerequisite pollutants are included in the control scope. Considering the impact of regional transmission, the pollution control of a certain place will include its adjacent areas in the control scope."

Lines 46-47: The reason of this long residence time is well known and must be added.

Reply:

- We revise it in the manuscript.
- "$PM_{2.5}$ in the atmosphere is either emitted directly or formed from gaseous precursors through complex gas phase, aqueous phase or heterogeneous chemical reactions (Cheng et al., 2016; Gao et al., 2016). $PM_{2.5}$ consists of complex composition, including sulfate, nitrate, organic carbon, elemental carbon, soil dust, and sea salt (Gao et al., 2016). It can stay in the boundary layer for a few days and in the free troposphere for a few weeks (similar to ozone). It can be efficiently removed out of atmosphere by precipitation, which is the major atmospheric sink (Jacob and Winner, 2009). It is mainly affected by the sources, meteorology, composition, transport and mixing as well as wet and dry removal rates of the associated $PM_{2.5}$ and the $PM_{2.5}$ exhibit substantial spatiotemporal variations(Jia and Jia, 2014; Poet et al., 1972)."

Lines 47-51: Not well explained, needs revision.

Reply:

- We revise it in the manuscript.
- "Research shows that $PM_{2.5}$ can stay in the atmosphere for several days to dozens of days, mainly affected by the sources, meteorology, composition, transport and mixing as well as wet and dry removal rates of the associated $PM_{2.5}$ and the $PM_{2.5}$ exhibit substantial spatiotemporal variations(Jia and Jia, 2014; Poet et al., 1972). Because of the sources and meteorology transport, the diurnal variation of $PM_{2.5}$ can range from several μg m^{-3} to many hundreds μg m^{-3} within several hours, and appreciable differences in $PM_{2.5}$ concentrations can occur within several kilometers spatially(Guo et al., 2017; Gupta and Christopher, 2009; Zheng et al., 2015). "

Line 69-70: Unclear for people not using those techniques.

Reply:

- We revise it in the manuscript. In the atmospheric physical and chemical model, the variation method is usually used to assimilate the observed data, and the objective results are obtained by optimizing covariance.
- "A Gridpoint Statistical Interpolation (GSI) Three-Dimensional Variational (3DVAR) data assimilation system is usually used to link the changes in AOD to aerosol chemical compositions (Gao et al., 2017; Saide et al., 2012)."

Lines 75-76: It would be interesting to know the reason of such low temporal resolution.

- Reply:

- Due to observation limitations such as clouds, the AOD observed by satellite is lacking under cloud and night. Previous studies mostly focused on mining the relationship between the satellite AOD and ground PM2.5 at the same spatial grid point, without fully mining the spatio-temporal correlation characteristics between the data, so there is a space gap, and most of them are filled in space through the accumulation of time.
- Also, the polar orbiting satellites are mostly used (MODIS AOD is the most commonly used data). Compared with the hourly resolution of geostationary satellites, their monitoring results are only once a day.

Line 78. Rephrase "…Yu et al., 2017). In addition, errors in this method can derive from the uncertainties…"

Reply:

- We revise it in the manuscript.
- "(Bi et al., 2019; Fang et al., 2016; He and Huang, 2018; Li et al., 2017; Ma et al., 2016; Park et al., 2020; Shtein et al., 2019; Wei et al., 2019; Wei et al., 2020; Xiao et al., 2017; You et al., 2016; Yu et al., 2017). In addition, errors in this method can derive from the uncertainties in chemical transport modeling or cloudy/hazy conditions."

Line 80: Change "offered" to "obtained". What do you mean by "hourly predictions of daytime PM2.5"?

Reply:

- We revise it in the manuscript. The satellite remote sensing aerosol optical thickness used for assimilation is usually passive remote sensing, and the night observation results cannot be obtained. For geostationary satellites, only daytime hourly AOD monitoring results can be obtained.
- "Due to the passive remote sensing monitoring, geostationary orbit satellites can only monitor the daytime aerosol optical thickness, and the related research is mainly about the daytime $PM_{2.5}$ concentration."

Lines 87-89: Statistical analyses on what? And why this result?

Reply:

- The statistics are based on the retrieval results of aerosol optical thickness of MODIS satellite. The MODIS satellite has strictly controlled its quality and removed clouds, ice, snow and serious haze. Therefore, the satellite AOD has a spatial lacking.

Line 90: Add "often impacted by" after "regions" (remove "with").

Reply:

- We revise it in the manuscript.
- "Better methods are thus needed to overcome these limitations, particularly for regions often impacted by thick clouds and severe haze pollution."

Line 93: Change “elements” to “variables”. And explain better whether “geographical information” is just position or also other details such as topography, land use, or other..

Reply:

- We revise it in the manuscript. We added the details of the input data. And we move Table 6 to the manuscript and added Table S3.
- “We use some geographic information data to explore its potential impact on $PM_{2.5}$. These inputs to the neural network include land cover types (MODIS land cover product at 0.05°×0.05° resolution, yearly), Normalized Differnce Vegetation Index (MODIS, 0.05°×0.05°, monthly), Enhanced Vegetation Index (MODIS, 0.05°×0.05°, monthly), road network (originally meter level, www.openstreetmap.org; Last access: July 10, 2020), point of interest data (POI), elevation data (1km×1km, the Resource and Environment Science Data Center, RESDC, http://www.resdc.cn; Last access: January 1, 2021), population/gross domestic product data from RESDC. These (road network, POI, elevation, etc.) were regridded to 0.01° grids. And they have the same grid center.”

Table 6. Descriptions of considered variables.

Product	Unit	Variable Definition	Spatial Resolution	Temporal Resolution
AOD		Aerosol optical depth	0.05°×0.05	1hour
Tempc	°C	Temperature	0.05°×0.05°× 12L	1hour
RH	%	Relative Humidity	0.05°×0.05°× 12L	1hour
HPBL	m	Planetary Boundary Layer Height	0.05°×0.05°	1hour
P	Hpa	Pressure	0.05°×0.05°× 12L	1hour
U	m/s	Wind Speed (U)	0.05°×0.05°× 12L	1hour
V	m/s	Wind Speed (V)	0.05°×0.05°× 12L	1hour
DEM	m	Digital Elevation Model	0.01°×0.01°	Annual
POI		Point of Interest	0.01°×0.01°	Annual
Traffic Network		Traffic Network	0.01°×0.01°	Annual
GDP	￥/km2	Gross Domestic Product	0.01°×0.01°	Annual
TPOP	people/km2	population density	0.01°×0.01°	Annual
Land Cover Type		Land Cover Type	0.05°×0.05°	Annual
EVI		Enhanced Vegetation Index	0.05°×0.05°	Monthly
NDVI		Normalized Differnce Vegetation Index	0.05°×0.05°	Monthly

Table S3. The input data shape

Category	Name	shape type	shape
AOD data	Himawari-8 Current	width,length,time	32,32,4
	Himawari-8 Closeness	width,length,time	32,32,10
	Himawari-8 Period	width,length,time	32,32,7
	MODIS	width,length,band×time	32,32,3×7
Meteorology	rh	width,length,time	32,32,9
	temperature	width,length,time	32,32,9
	pressure	width,length,time	32,32,9
	hpbl	width,length,time	32,32,9
	u	width,length,time	32,32,9
	v	width,length,time	32,32,9
	rh	width,length,height	32,32,12
	temperature	width,length,height	32,32,12
	pressure	width,length,height	32,32,12
	hpbl	width,length,height	32,32,1
	u	width,length,height	32,32,12
	v	width,length,height	32,32,12
Geographic information data	POI	width,length,type	64,64,7
	Traffic Network	width,length,type	64,64,9
	DEM	width,length,type	64,64,1
	GDP	width,length,type	64,64,1
	Tpop	width,length,type	64,64,1
	Land Cover Type	width,length,type	32,32,17
	EVI	width,length,type	32,32,1
	NDVI	width,length,type	32,32,1

Lines 93-95: This type of statements are not appropriate for the Introduction section. Better to replace it with information on the article structure.

Reply:

- We revise it in the manuscript. We modify this section to introduce the structure of the article.
- "In section two, we introduce the materials and methods and in section three we description the results, with the ST-NN model, we get the high spatial and temporal resolution (0.01° and hourly) surface $PM_{2.5}$ even during nighttime and under cloudy or hazy conditions. And we have carried on the detailed verification and discussion to the result."

Lines 97-99: Please add more details on the regions (e.g., better description, maps). This sentence is not that useful in the current version. Also you can introduce the different subsections that the reader will encounter in this section.

Reply:

- In the updated manuscript. We revised this section to introduce the time and space scope of the study and the overview of this section.
- "We built a deep learning model to improve estimates of ground-level $PM_{2.5}$ concentrations, particularly for regions without sampling sites, and for conditions (cloudy, hazy, nighttime, etc.) where satellite retrievals are not available. Our research area is concentrated in the Middle East of China (Figure. S3), with a time span of four years from 2017 to 2020. In Section 2.1, we introduced the input data of the model, in Section 2.2 and 2.3, we introduced the model structure and training verification methods, and in Section 2.4, we introduced the sensitivity analysis method used to quantitatively analyze the impact of the input data on the results. This method is used to open the black box model of neural network."

Lines 101-104: Please provide references on the data sources (e.g.. websites).

Reply:

- We revise it in the manuscript. We added the source of the data.
- "We used hourly ground-level observations of $PM_{2.5}$ from the Chinese National Environmental Monitoring Center (CNMEC) network (http://www.cnemc.cn/), the daily MODerate Resolution Imaging Spectroradiometer (MODIS) 3km aerosol products (Levy et al., 2013) (https://doi.org/10.5067/MODIS/MOD04_3K.061), and the hourly 0.05°×0.05° Himawari-8 AOD products(Yoshida et al., 2018) (https://doi.org/10.2151/jmsj.2018-039)."

Line 108: Until now, you have always referred to AOD. Please explain what is aerosol optical thickness (AOT) and its relation with AOD.

Reply:

- Aerosol optical depth (AOD) and aerosol optical thickness (AOT) mean the same.

Lines 107-110: References needed.

Reply:

- In the updated manuscript. We have added corresponding references.
- "The reason for using two different types of satellite aerosol optical thickness is that the Himawari-8 satellite aerosol is effective in capturing the daily variation of aerosols (Tan et al., 2022), while the MODIS aerosol product has aerosol optical thickness in different bands to capture information on the properties of aerosols and more accurate numerical results (Levy et al., 2013)."
- Tan, Y., Wang, Q., and Zhang, Z.: Assessing spatiotemporal variations of AOD in Japan based on Himawari-8 L3 V31 aerosol products: Validations and applications, Atmospheric Pollution Research, 13, 101439, https://doi.org/10.1016/j.apr.2022.101439, 2022.
- Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., and Hsu, N.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 2013

Line 119: Which meteorological fields? Of which variables?

Lines 121-122: Validation against what? Please provide a discussion on the validation results.

Reply:

- In the updated manuscript. We added a detailed introduction to the meteorological boundary field, and the meteorological field input by our final model. Table 6 and Table S3 show the input dataset. In order to evaluate the validation results of WRF model and ground station, validation was carried out.
- "The boundary field parameters at 26 mandatory levels from 1000 millibars to 10 millibars. The boundary field parameters include surface pressure, sea level pressure, geopotential height, temperature, sea surface temperature, soil values, ice cover, relative humidity, u- and v- winds, vertical motion, vorticity and ozone."
- "In order to evaluate the validation results of WRF model and ground station, validation was carried out. Figure S1 shows the verification results of main meteorological variables (monitoring of coincidence between ground meteorological stations and demand variables). The R^2 of temperature and pressure is above 0.95, and the slope is close to 1. The R of wind speed is greater than 0.65, which is mainly because the wind speed has a strong local variability, which makes the difference between the analog quantity (local average) and the observation (fixed point)."

Table 6. Descriptions of considered variables.

Product	Unit	Variable Definition	Spatial Resolution	Temporal Resolution
AOD		Aerosol optical depth	0.05°×0.05	1hour
Tempc	°C	Temperature	0.05°×0.05°×12L	1hour
RH	%	Relative Humidity	0.05°×0.05°×12L	1hour
HPBL	m	Planetary Boundary Layer Height	0.05°×0.05°	1hour
P	Hpa	Pressure	0.05°×0.05°×12L	1hour
U	m/s	Wind Speed (U)	0.05°×0.05°×12L	1hour
V	m/s	Wind Speed (V)	0.05°×0.05°×12L	1hour
DEM	m	Digital Elevation Model	0.01°×0.01°	Annual
POI		Point of Interest	0.01°×0.01°	Annual
Traffic Network		Traffic Network	0.01°×0.01°	Annual
GDP	￥/km2	Gross Domestic Product	0.01°×0.01°	Annual
TPOP	people/km2	population density	0.01°×0.01°	Annual

Land Cover Type	Land Cover Type	0.05°×0.05°	Annual
EVI	Enhanced Vegetation Index	0.05°×0.05°	Monthly
NDVI	Normalized Differnce Vegetation Index	0.05°×0.05°	Monthly

Table S3. The input data shape

Category	Name	shape type	shape
AOD data	Himawari-8 Current	width,length,time	32,32,4
	Himawari-8 Closeness	width,length,time	32,32,10
	Himawari-8 Period	width,length,time	32,32,7
	MODIS	width,length,band×time	32,32,3×7
Meteorology	rh	width,length,time	32,32,9
	temperature	width,length,time	32,32,9
	pressure	width,length,time	32,32,9
	hpbl	width,length,time	32,32,9
	u	width,length,time	32,32,9
	v	width,length,time	32,32,9
	rh	width,length,height	32,32,12
	temperature	width,length,height	32,32,12
	pressure	width,length,height	32,32,12
	hpbl	width,length,height	32,32,1
	u	width,length,height	32,32,12
	v	width,length,height	32,32,12
Geographic information data	POI	width,length,type	64,64,7
	Traffic Network	width,length,type	64,64,9
	DEM	width,length,type	64,64,1
	GDP	width,length,type	64,64,1
	Tpop	width,length,type	64,64,1
	Land Cover Type	width,length,type	32,32,17
	EVI	width,length,type	32,32,1
	NDVI	width,length,type	32,32,1

Line 122: Here and throughout the text, I don't understand the need to have a point between "Figure" and the figure number.

Reply:

- In the updated manuscript, we removed it.

Lines 127-128: Not clear.

Reply:

- In the updated manuscript, we rewrote this part.
- "The input data includes satellite AOD, meteorological and geographic

information data (Table S3 and Table 6). The central grid of all input data is the CNEMC site label location."

Table S3 The input data shape

Category	Name	shape type	shape
AOD data	Himawari-8 Current	width,length,time	32,32,4
	Himawari-8 Closeness	width,length,time	32,32,10
	Himawari-8 Period	width,length,time	32,32,7
	MODIS	width,length,band×time	32,32,3×7
Meteorology	rh	width,length,time	32,32,9
	temperature	width,length,time	32,32,9
	pressure	width,length,time	32,32,9
	hpbl	width,length,time	32,32,9
	u	width,length,time	32,32,9
	v	width,length,time	32,32,9
	rh	width,length,height	32,32,12
	temperature	width,length,height	32,32,12
	pressure	width,length,height	32,32,12
	hpbl	width,length,height	32,32,1
	u	width,length,height	32,32,12
	v	width,length,height	32,32,12
Geographic information data	POI	width,length,type	64,64,7
	Traffic Network	width,length,type	64,64,9
	DEM	width,length,type	64,64,1
	GDP	width,length,type	64,64,1
	Tpop	width,length,type	64,64,1
	Land Cover Typr	width,length,type	32,32,17
	EVI	width,length,type	32,32,1
	NDVI	width,length,type	32,32,1

Table 6. Descriptions of considered variables.

Product	Unit	Variable Definition	Spatial Resolution	Temporal Resolution
AOD		Aerosol optical depth	0.05°×0.05	1hour
Tempc	°C	Temperature	0.05°×0.05°×12L	1hour
RH	%	Relative Humidity	0.05°×0.05°×12L	1hour
HPBL	m	Planetary Boundary Layer Height	0.05°×0.05°	1hour
P	Hpa	Pressure	0.05°×0.05°×12L	1hour
U	m/s	Wind Speed (U)	0.05°×0.05°×	1hour

V	m/s	Wind Speed (V)	12L 0.05°×0.05°× 12L	1hour
DEM	m	Digital Elevation Model	0.01°×0.01°	Annual
POI		Point of Interest	0.01°×0.01°	Annual
Traffic Network		Traffic Network	0.01°×0.01°	Annual
GDP	￥/km2	Gross Domestic Product	0.01°×0.01°	Annual
TPOP	people/km2	population density	0.01°×0.01°	Annual
Land Cover Type		Land Cover Type	0.05°×0.05°	Annual
EVI		Enhanced Vegetation Index	0.05°×0.05°	Monthly
NDVI		Normalized Differnce Vegetation Index	0.05°×0.05°	Monthly

Lines 134-135: Not clear.

Reply:

- In the updated manuscript, we rewrote this part.
- "Pearson correlation coefficient can test whether two continuous variables have potential correlation in statistics. For meteorological variables and satellite data, we use Pearson significance level to judge whether they and target variables are significant, and determine if they are used as model inputs. Chi-square test is a commonly used method to test the statistical correlation between discrete variables. For geographic information variables, they are approximately considered unchanged during our research. First, we use k-mean method to cluster them for reducing dimensions. We classify the annual average data of CNEMC $PM_{2.5}$ according to WHO Global Air Quality Guidelines (<10μg m^{-3}, 10~15μg m^{-3}, 15~25μg m^{-3}, 25~35μg m^{-3}, 35~50μg m^{-3}, 50~75μg m^{-3}, 75~100μg m^{-3}, >100μg m^{-3}) (WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Executive summary). Then we use Chi-square test to test whether it has statistical correlation. Only those parameters that pass the significance test (α < 0.01) were selected (Table S7)."

Lines 135-137: How did you classify the severity of pollution?

Reply:

- In the updated manuscript, we rewrote this part. We classified ground $PM_{2.5}$ according to WHO guidelines (<10μg m^{-3}, 10~15μg m^{-3}, 15~25μg m^{-3}, 25~35μg m^{-3}, 35~50μg m^{-3}, 50~75μg m^{-3}, 75~100μg m^{-3}, >100μg m^{-3}).
- "We use the annual average CNEMC $PM_{2.5}$ concentration data and classify it (<10μg m^{-3}, 10~15μg m^{-3}, 15~25μg m^{-3}, 25~35μg m^{-3}, 35~50μg m^{-3}, 50~75μg m^{-3}, 75~100μg m^{-3}, >100μg m^{-3}) according to the WHO Global Air Quality Guidelines

(WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Executive summary)."

Lines 125-159: This section is very unclear. Several methods are listed, but I cannot understand how most of them were used.

Reply:

- In the revised manuscript, we have revised this part, and described it according to data introduction, data selection, model structure, model training and testing, and sensitivity analysis, also we introduced specific methods for each part.
- "

  2.1.2 Datasets Selection

  To select the potentially relevant data to the target variable, We need to filter the data to reduce the complexity of the model, so as to obtain the expected results in a limited time. First, we conduct data preselection, and select data by checking the correlation between the input data variables and our target output (ground $PM_{2.5}$ concentration). All input features were used to examine the potential relationship with $PM_{2.5}$.

  Pearson correlation coefficient can test whether two continuous variables have potential correlation in statistics. For meteorological variables and satellite data, we use Pearson significance level to judge whether they and target variables are significant, and determine if they are used as model inputs. Chi-square test is a commonly used method to test the statistical correlation between discrete variables. For geographic information variables, they are approximately considered unchanged during our research. First, we use k-mean method to cluster them for reducing dimensions. We classify the annual average data of CNEMC $PM_{2.5}$ according to WHO Global Air Quality Guidelines (<10μg m^{-3}, 10~15μg m^{-3}, 15~25μg m^{-3}, 25~35μg m^{-3}, 35~50μg m^{-3}, 50~75μg m^{-3}, 75~100μg m^{-3}, >100μg m^{-3}) (WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Executive summary). Then we use Chi-square test to test whether it has statistical correlation. Only those parameters that pass the significance test ($\alpha < 0.01$) were selected (Table S7).

  2.1.3 Datasets Filter

  For CNEMC $PM_{2.5}$ data, we remove null data. Secondly, we also tested the performance of the model under different data filtering rules (Table S10).

  All hourly data at a specific monitoring site were transformed into z scores, and then the transformed data (Zi) were removed if they met one of the following conditions: (1) the absolute Zi was larger than 4 ($|Z_i|>4$), (2) the increment of Z_i from the previous hourly value was larger than 9 ($|Z_i-Z_{i-1}|>9$), or (3) the ratio of the z score to its centered moving average of order 3 (MA3) was larger than 2 ($3Z_i/(Z_{i-1}+Z_i+Z_{i+1})>2$) (Su et al., 2019). For missing satellite AOD data, we use null values.

2.2 Data Preprocessing and ST-NN Model Configuration

Figure S2 displays the architecture of the deep learning ST-NN model. The input data includes satellite AOD, meteorological and geographic information data (Table 6 and Table S3). The central grid of all input data is the CNEMC site label location.

For the data of UTC 4:00~09:00, we use the previous four hours Himawari-8 AOD, the past days Himawari-8 AOD (daytime AOD of the past two days) and the past week Himawari-8 AOD (daily average) to formulate the influencing factors. For other times, we use the past days Himawari-8 AOD (daytime AOD of the past two days) and the past week Himawari-8 AOD (daily average) to formulate the influencing factors. MODIS AOD values retrieved at three bands were used for the past week's results. First, feature extraction (nonlinear transformation) is performed on each data layer, and then data fusion is performed. The data fusion process can be expressed as：

$$Z_{fusion} = \sum_{i=1}^{c} K_i * X_{AOD_{himawari-8_{current}}} + \sum_{i=1}^{c} K_i * X_{AOD_{himawari-8_{closeness}}} + \sum_{i=1}^{c} K_i * X_{AOD_{himawari-8\ period}} + \sum_{i=1}^{c} K_i * X_{AOD_{MODIS}}$$

Where ＊ is convolution and K_i is learnable parameter.

Meteorological data were arranged as time series of the bottom model level and the vertical features at the current moment t_0 to include the influences of temporal and spatial evolution. For meteorological data and geographic information data, we use the same method to extract the data features and then perform data fusion. For data with different spatial resolutions, we use the upper sampling layer in the feature extraction process to make them have the same spatial size before data fusion.

We mainly use the Inception-Resnet block and pooling layers for feature extraction, which has been proved to be able to quickly and effectively mine the potential features of multidimensional data(more details in Figure S2) (Szegedy et al., 2017). And $PM_{2.5}$ concentrations were then obtained through optimizing the Log-Cosh loss function below.

$$Loss = \frac{1}{n}\sum_{i=1}^{n} \log\left(\cosh\left(y_i^{predict} - y_i^{true}\right)\right) \qquad (1)$$

where $y_i^{predict}$ means the model predicted value and the y_i^{true} represents the observations.

For small differences the log-cosh loss performances similar as $\frac{(y_i^{predict}-y_i^{true})^2}{2}$, and for huge differences (at the beginning of model training), it's closer abs($y_i^{predict} - y_i^{true}$)-log(2).

We initialized all the layers with the built-in Keras glorot uniform initializer as 0, and the biases were initialized with 0. Due to the symmetry of the data, tanh was used as the activation function. We trained the networks for 64 epochs with a batch size of 4, and SGD (Stochastic Gradient Descent) optimizer with an exponential decay of the learning rate α as:

$$\alpha = \begin{cases} 0.001 & epoch <= 32 \\ 0.001 \times \exp(0.1 \times (32\text{-epoch})) & epoch > 32 \end{cases} \quad (2)$$

"

Lines 161-163: So you used just 10% of data for testing: isn't this test period too short?

Reply:

In our study, we used cross validation, which means that all the data were used as testing.

- "10-CV has been proved to be a reasonable means to evaluate results, such as Table References cited in Table. S9. And, we have a large dataset to avoid over fitting.
- Compared with other studies, our hourly full coverage ground PM2.5 concentration prediction has greatly increased the amount of data. The sample size exceeds 1 million in each study area and year (Table. S11 and S12). Also we compared the results under different training and test proportions, and showed that 10-CV was reasonable in this study (Table.S16)."
- And we used the Beijing site for independent verification. The following figure shows the verification results. The black points are the data of CNEMC sites participating in the training, and the red points are the testing data of Beijing control sites. The R^2 is above 0.86 and the RMSE is less than 24 μg·m^{-3}.

Table S16. Validation results under different training and test proportions

test:train	1:9	2:8	3:7	4:6	5:5	6:4	7:3
R-squre	0.81	0.81	0.78	0.77	0.76	0.76	0.74
RMSE(μg/m³)	16.15	16.6	17.57	17.87	18.67	18.55	19.32

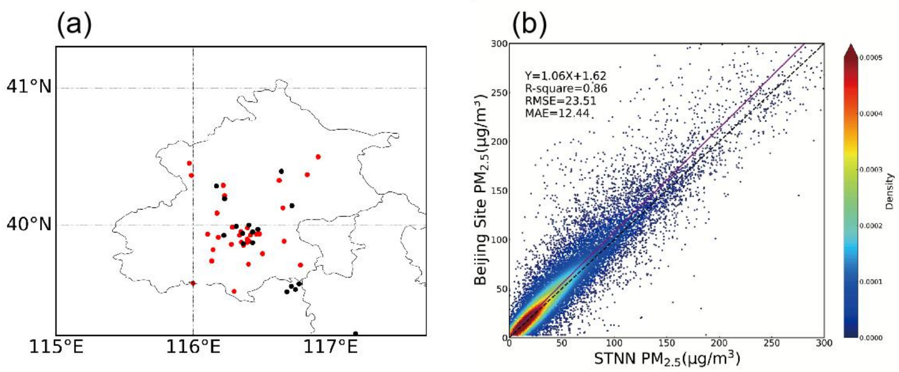

Figure. Review. (a) Distribution of verified Beijing control sites (red) and CNEMC sites (black). (b) shows the verification results of the ST-NN model and Beijing sites.

Lines 169-170: References to the libraries are needed.

Reply:

- In the updated manuscript, we added the source of using the software library.
- "The proposed model was implemented in Python 3.7 (https://www.python.org/) with a neural network library named Keras (https://keras.io/guides/) and

TensorFlow as the backend (https://www.tensorflow.org/).”

Lines 161-170: Also this section is quite unclear, without references and details that can help the reader to repeat the process if needed.

Reply:

- In the revised manuscript, we added references. Readers can reproduce our experimental process according to the data and methods introduced in this section.
- “Based on number of samples, dimension of time and dimension of space, the entire datasets (one year) were randomly sorted into 10 sections, with 9 sections for training and the rest for testing(Schultz et al., 2021). 10-CV has been proved to be a reasonable means to evaluate results, such as Table References cited in Table. S9. And, we have a large dataset to avoid over fitting. Compared with other studies, our hourly full coverage ground PM2.5 concentration prediction has greatly increased the amount of data. The sample size exceeds 1 million in each study area and year (Table. S11 and S12). Also we compared the results under different training and test proportions, and showed that 10-CV was reasonable in this study (Table.S16)
- The shape of the input dataset is shown on Figure.S2. For sample-based cross-validation, we randomly grouped all the data; for spatial-based cross-validation, we randomly grouped the data by site location; and for temporal-based cross-validation, we randomly grouped the data by time. The testing data doesn’t participate in the model training process. The proposed model was implemented in Python 3.7 (https://www.python.org/) with a neural network library named Keras (https://keras.io/guides/) and TensorFlow as the backend (https://www.tensorflow.org/).”

Lines 172-204: The aim of this section is quite obscure.

Reply:

- Usually in black box models, it is difficult to quantify the impact of input variables.
- We open the neural network black box model by sensitivity analysis, and quantitatively analyze the impact of each input variable on the results. These analysis indexes are used to measure the impact of each input variable on the results. This section explains the sensitivity analysis method we adopted.
- In the updated version, we first introduced the purpose of this part.
- “The neural network model is generally considered as a black box model, and we open the black box model through sensitivity analysis and visualization to analyze the quantitative impact of each input variable on the ground $PM_{2.5}$ concentration. (Cortez and Embrechts, 2013).”

Lines 209-210:Well, not really, as the methodology section lacks many details for instance on the kind of meteorological data used.

Reply:

- In the updated manuscript, we rewrote the method section. Added a detailed description of the input data and tables. From Figure S2 and Table S3, the input data of the model can be clearly obtained.

- “2.1.1 Datasets Description
- We used hourly ground-level observations of $PM_{2.5}$ from the Chinese National Environmental Monitoring Center (CNMEC) network(http://www.cnemc.cn/) as the label and the validation data, they use β X-ray method and vibration balance method for measurement.
- For the input datasets, the daily MODerate Resolution Imaging Spectroradiometer (MODIS) 3km aerosol products(Levy et al., 2013a)(https://doi.org/10.5067/MODIS/MOD04_3K.061), and the hourly 0.05°×0.05° Himawari-8 AOD products (Yoshida et al., 2018) (https://doi.org/10.2151/jmsj.2018-039) are used.
- The original MODIS products were mapped onto regional grids of 0.05°×0.05° resolution to ensure that the input data have the same grid center. Considering the diurnal variation of the solar zenith angles, only daytime satellite data (defined as 00:00-09:00 UTC) were used in this study. The reason for using two different types of satellite aerosol optical thickness is that the Himawari-8 satellite aerosol is effective in capturing the daily variation of aerosols(Tan et al., 2022), while the MODIS aerosol product has aerosol optical thickness in different bands to capture information on the properties of aerosols and more accurate numerical results(Levy et al., 2013b).
- We use some geographic information data to explore its potential impact on $PM_{2.5}$. These inputs to the neural network include land cover types (MODIS land cover product at 0.05°×0.05° resolution, yearly), Normalized Differnce Vegetation Index (MODIS, 0.05°×0.05°, monthly), Enhanced Vegetation Index (MODIS, 0.05°×0.05°, monthly), road network (originally meter level, www.openstreetmap.org; Last access: July 10, 2020), point of interest data (POI), elevation data (1km×1km, the Resource and Environment Science Data Center, RESDC, http://www.resdc.cn; Last access: January 1, 2021), population/gross domestic product data from RESDC. These (road network, POI, elevation, etc.) were regridded to 0.01° grids. And they have the same grid center.
- Meteorological variables are important factors affecting atmospheric aerosols. The weather fields (0.05° × 0.05°, hourly) simulated by the Weather Research and Forecasting (WRF) model version 4.0 with three nested domains (https://www.mmm.ucar.edu/weather-research-and-forecasting-model; Last access: May 15, 2020). The initial conditions and boundary conditions of the meteorological fields were derived from the National Centers for Environment Prediction’s (NCEP) 6-hour final operational global (FNL) data with a spatial resolution of 0.25° × 0.25°. The boundary field parameters at 26 mandatory levels from 1000 millibars to 10 millibars. The boundary field parameters include surface pressure, sea level pressure, geopotential height, temperature, sea surface temperature, soil values, ice cover, relative humidity, u- and v- winds, vertical motion, vorticity and ozone. And the parameters we used is shown in Table 6 and Table S3. In order to evaluate the validation results of WRF model and ground

station, validation was carried out. Figure. S1 shows the verification results of main meteorological variables (monitoring of coincidence between ground meteorological stations and demand variables). The R^2 of temperature and pressure is above 0.95, and the slope is close to 1. The R of wind speed is greater than 0.65, which is mainly because the wind speed has a strong local variability, which makes the difference between the analog quantity (local average) and the observation (fixed point). Descriptions and features of these considered datasets are listed in Table 6 and Table S4-S6. ”

“

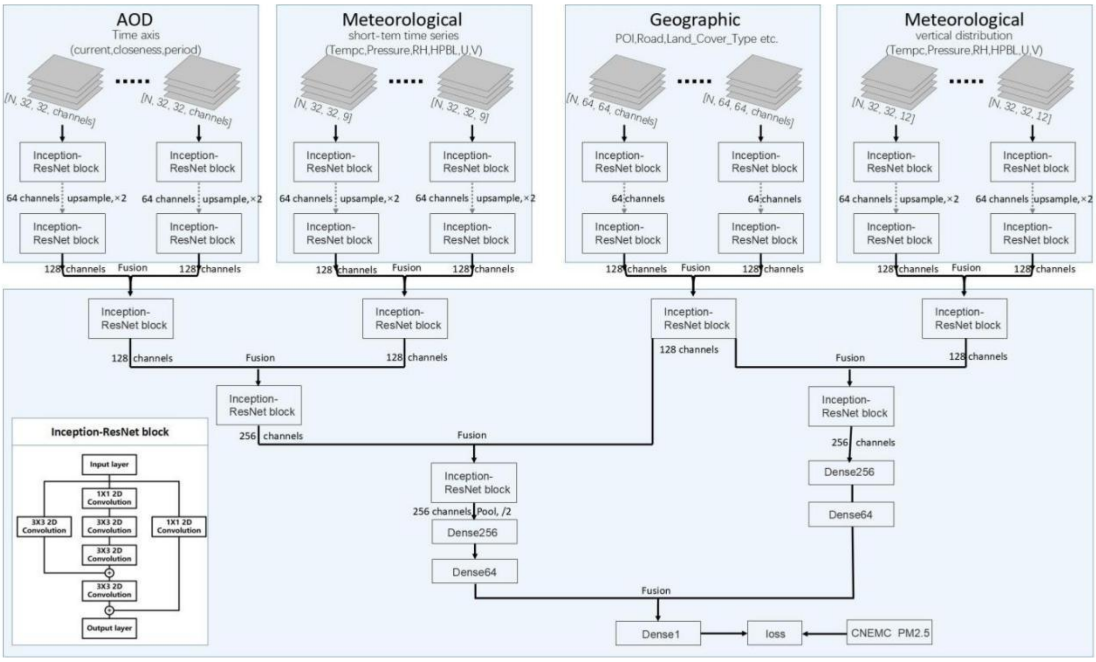

Figure. S2. The architecture of the ST-NN model. Inputs data include AOD, meteorological data and geographic data. All the input variables have the 4-D dimensions as [N, lat_size, lon_size, channel]. N means the batch size, lat_size and lon_size means the scale of the data at latitude and longitude, channels represent the types or the height/time dimension of the data. Considering the computational efficiency and the generalization ability of the model, we choose N as 4. First, feature extraction was used for individual data by Inception-ResNet, which is an efficient feature extraction process. After it, we up-sampled the data at 0.05°×0.05°resolution using the transposed convolution layer, which is a learning-based up-sampling method. We used the strides as 2 and a 2×2 convolution kernel to double size of the input data. Accordingly, all data had the same size. Then we mined the characteristics of each variable and fused the data with same types by concatenate layer. Then the temporal and spatial features were extracted by fusing the time-series of aerosol and meteorological data with the geographic information data. The final result was obtained through the fully connected layer.

Table S3. The input data shape

Category	Name	shape type	shape
AOD data	Himawari-8 Current	width,length,time	32,32,4

	Himawari-8 Closeness	width,length,time	32,32,10
	Himawari-8 Period	width,length,time	32,32,7
	MODIS	width,length,band×time	32,32,3×7
Meteorology	rh	width,length,time	32,32,9
	temperature	width,length,time	32,32,9
	pressure	width,length,time	32,32,9
	hpbl	width,length,time	32,32,9
	u	width,length,time	32,32,9
	v	width,length,time	32,32,9
	rh	width,length,height	32,32,12
	temperature	width,length,height	32,32,12
	pressure	width,length,height	32,32,12
	hpbl	width,length,height	32,32,1
	u	width,length,height	32,32,12
	v	width,length,height	32,32,12
Geographic information data	POI	width,length,type	64,64,7
	Traffic Network	width,length,type	64,64,9
	DEM	width,length,type	64,64,1
	GDP	width,length,type	64,64,1
	Tpop	width,length,type	64,64,1
	Land Cover Type	width,length,type	32,32,17
	EVI	width,length,type	32,32,1
	NDVI	width,length,type	32,32,1

"

Line 211: Change "that" to "when".

Reply:

- We revise it in the manuscript.
- "particularly for regions without sampling sites, and for conditions (cloudy, hazy, nighttime, etc.) when satellite retrievals are not available."

Line 216: What about the R2 or R value?

Reply:

- In the updated manuscript, we added the results of R^2.
- "As displayed in Figure. 1, our ST-NN model accurately captured the observed spatiotemporal variability of daytime $PM_{2.5}$, with regression slopes close to 1 and intercepts close to 0. And the R^2 is above 0.8."

Lines 225: Change "pleasant" to "good".

Reply:

- In the updated manuscript, we revise it.
- "Its applications to other years and over other regions in China achieved similar good performance"

Lines 222-226: What about the other parameters (RMSE, MAE, …)?

- In the updated manuscript, we added and discussed the results of verifying RMSE, MAE and Solp.
- "R^2 values with respect to sampling selection, temporal variability and spatial distribution in North China were generally above 0.85, RMSE were less than 26 μgm^{-3} and MAE were were less than 16 μgm^{-3} (Figure. 1), indicating the applicability of the model under various complex conditions. Its applications to other years and over other regions in China achieved similar good performance, and R^2 value reached even 0.90 when it is applied to Shaanxi Province for year 2019 (Table 1). And the RMSE were less than 23 μgm^{-3} (Table 2-1). MAE were less than 16 μgm^{-3} (Table 2-2). Slop close to 1 (Table 2-3).

Table 2-1. RMSE of cross validation with respect to spatial distribution.

	2017		2018		2019		2020	
	day	night	day	night	day	night	day	night
North China	19.77	22.59	19.92	19.86	16.53	18.44	16.46	13.99
East China	16.15	16.51	13.09	14.04	13.19	12.13	9.88	9.47
South China	11.11	12.81	10.38	11.38	9.52	11.41	6.00	8.96
Sichuan Basin	14.80	17.52	13.90	18.51	10.28	11.86	8.03	10.74
Shaanxi Province	20.15	22.79	15.47	18.88	15.14	17.13	12.01	12.33

Table 2-2. MAE values of cross validation of the model with respect to spatial distribution.

	2017		2018		2019		2020	
	day	night	day	night	day	night	day	night
North China	11.75	15.02	11.91	12.15	8.94	10.86	8.93	8.24
East China	9.54	10.41	8.68	9.27	8.76	8.17	6.94	6.51
South China	6.82	7.84	6.77	8.22	6.53	7.34	4.14	6.01
Sichuan Basin	9.27	10.32	9.41	11.37	6.69	8.04	5.74	7.47
Shaanxi Province	12.48	13.96	9.98	11.97	9.15	10.48	7.80	8.03

Table 2-3. Slop values of cross validation of the model with respect to spatial distribution.

	2017		2018		2019		2020	
	day	night	day	night	day	night	day	night
North China	1.03	0.97	0.99	0.99	0.99	0.98	0.99	1.04
East China	1.07	1.00	0.97	1.05	0.96	1.00	0.98	1.04
South China	1.02	1.06	1.03	1.00	1.00	1.01	0.98	0.96
Sichuan Basin	1.01	1.00	1.01	1.03	1.03	1.04	1.04	1.02
Shaanxi Province	1.02	1.03	1.01	1.00	0.98	0.99	1.02	0.98

"

Lines 210-212: If you do not have sampling sites and satellite retrievals, how can you train the model and how can you test (I mean, which data can you use as “measured PM2.5” in Figure 1)?

Reply:

For satellite data, the input isn’t a point, but a 3D spatiotemporal data. We use neural network model to mine the spatiotemporal relationship of data to obtain the ground $PM_{2.5}$ concentration.

Our model is a neural network based on data, so we need sites as labels for training. At present, there are more than 1600 state-controlled stations in China. The distribution is shown in Figure. S13, which is representative in space. Moreover, we use random sampling to make our test process representative of different spatial features.

In addition, we also verified our results through the Beijing control site (Independent of CNEMC) (Figure.S5 and Figure. Review.). The following figure shows the verification results. The black points are the data of CNEMC sites participating in the training, and the red points are the testing data of Beijing control sites. The R^2 is above 0.86 and the RMSE is less than 24 μg·m^{-3}.

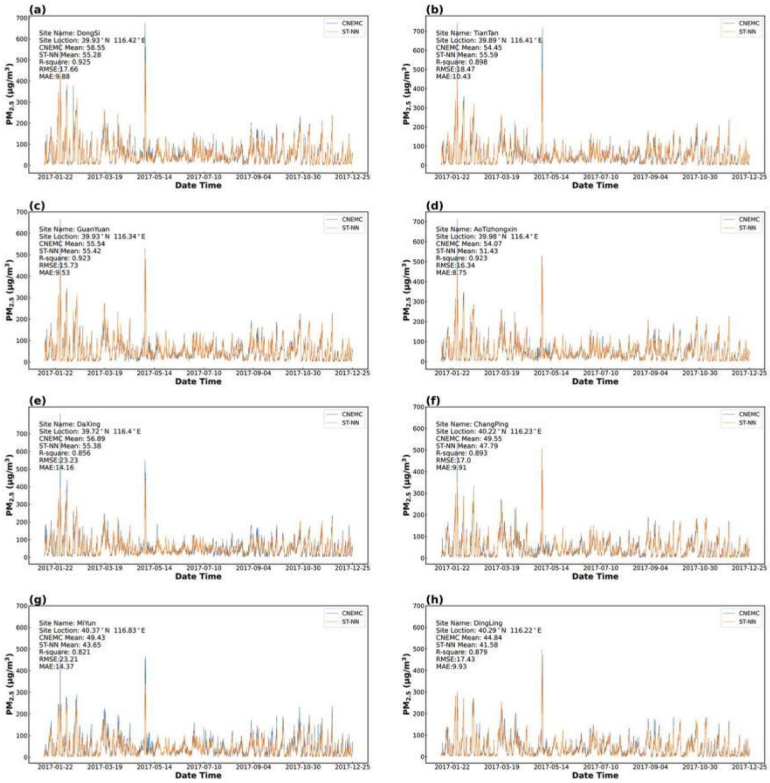

Figure. S5. ST-NN model predicted and ground-level observed (not used in training) time series of PM2.5 in Beijing stations. (a) Dongsi station in Beijing. (a) Tiantan station in Beijing. (c) Guanyuan station in Beijing. (d) Aotizhognxin station in Beijing. (e) Daxing station in Beijing. (f) Changping station in Beijing. (g) Miyun station in Beijing. (h) Dingling station in Beijing. (a-d) stations are in city, and (e-h) are rural stations.

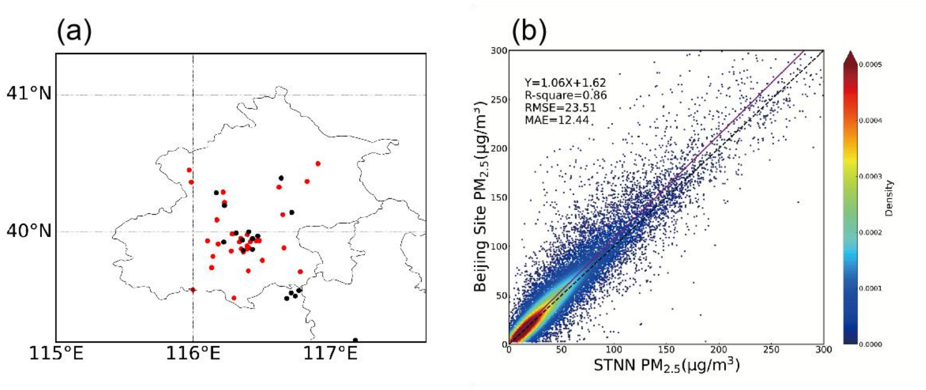

Figure. Review. (a) Distribution of verified Beijing control sites (red) and CNEMC sites (black). (b) shows the verification results of the ST-NN model and Beijing sites.

Line 227: Not all aerosols survive that long in the atmosphere! And also, the residence time is essentially driven by wet deposition processes. I assume that under monsoons period aerosols do not last long in the atmosphere..

Reply:

- Of course, the residence time of aerosols in the atmosphere has different time scales, especially in the monsoon and rainy periods, aerosols will settle quickly in the atmosphere. It mainly affected by the sources, meteorology, composition, transport and mixing as well as wet and dry removal rates of the associated $PM_{2.5}$ and the $PM_{2.5}$ exhibit substantial spatiotemporal variations(Jia and Jia, 2014; Poet et al., 1972). So we input meteorological and satellite data at different time scales, and hope to capture the characteristics of aerosol changes in the atmosphere by mining the potential relationship between multi-source data and near surface $PM_{2.5}$ concentration.
- In the updated manuscript, we have revised it.
- "Considering the potential spatialtemporal relationship between various variables and ground $PM_{2.5}$ under different conditions (Jia and Jia, 2014; Poet et al., 1972) and the lifetime of aerosol (Williams et al., 2002)."

Lines 231-232: This value is not that low..

Reply:

- There are some differences validation results in different time, and the lower limit of the data is selected as a conservative explanation.

Line 232: Please avoid the use of these non scientific terms ("delightful").

Reply:

- We revise it in the manuscript.
- "Similar performances were got for other regions also, and the performance of the model in predicting nighttime $PM_{2.5}$ did not exhibit a significant degradation from daytime (Table 1)."

Lines 242-247: Please discuss also other metrics apart from R^2.

Reply:

- In the revised manuscript, we discussed other parameters except R^2.
- "In addition to cross-validation, independent validation of this ST-NN model was conducted with $PM_{2.5}$ concentrations observed at sites that were not included in the model training. The variability of $PM_{2.5}$ concentrations at these independent stations were also accurately captured by our model, with R^2 values greater than 0.8, RMSE less than 24 μgm^{-3}, MAE less than 15 μgm^{-3} (Figure. S5). Independent validation was conducted also with respect to the diurnal variation of $PM_{2.5}$. As indicated in Figure. S6, the diurnal pattern of $PM_{2.5}$ over multiple independent stations across China was reproduced by the ST-NN model. The R^2 values greater than 0.8, RMSE less than 24 μgm^{-3}, MAE less than 14 μgm^{-3} (Figure. S6)."

Lines 281-282: Wet removal is not connected with presence of clouds but of precipitation.

Reply:

- In the updated manuscript, we revised the statement.
- "In cloudy scenes, $PM_{2.5}$ concentrations exhibited lower values when relative humidity (RH) > 60% in South China (Figure. S12c). This could be related to cloud-precipitation-related wet removal of air pollutants."

Lines 311-315: This seems a sort of repetition of the Introduction.

Reply:

- In the updated manuscript, we removed it.

Lines 306-361: The Discussion section seems just like a list of advantages of the methods, rather than a true discussion of the results. For instance, I failed to understand what are the variables that finally enter the model, and why the other variables are probably not affecting PM. Also, reasons for limitations, issues are discussed with small details. References against which to compare the results are given, but references on how to interpret the results are instead not given. Finally, a conclusion section is missing.

Reply:

- In the revised manuscript, we added more discussion about the results.
- The input of our model has been introduced in detail before, and here we discuss the limitations. Of course, we expect to get more related variables for input, such as some real-time emission data. But because we can only get limited observation data, we only predict the ground $PM_{2.5}$ concentration based on the data we have. In the future, we expect to get more and more comprehensive monitoring data to improve the accuracy of our result.
- We added the discussion and limitation of the results obtained, the reasons for the influence of each variable on the results and the analysis of the reasons for the difference errors under different pollution levels.
- In addition, compared with similar studies, we mine the spatiotemporal information of multi-source data to obtain the ground $PM_{2.5}$ concentration, which increases the spatiotemporal coverage of data.
- And we added a conclusion section.

- “A number of studies have explored the prediction of ground-level $PM_{2.5}$ concentrations with statistical methods, as indicated in Table S9. Despite that many of these studies achieved similar performance with respect to R^2, RMSE, MAE and slope as our study, these studies can not get the results of full coverage of time and space. (Fang et al., 2016; Ma et al., 2016; You et al., 2016; Li et al., 2017; Xiao et al., 2017; Yu et al., 2017; He and Huang, 2018; Bi et al., 2019; Shtein et al., 2019; Wei et al., 2019; Park et al., 2020; Wei et al., 2020). This is mainly because they mainly focus on the relationship between multi-source data seeking corresponding positions and ground $PM_{2.5}$, and do not fully mine the spatiotemporal relationship of data. In this study, we fully used the spatiotemporal features of aerosol and simulated the dynamic evolution of aerosols under complex influences of meteorology, terrain, etc. We select the input data through the correlation test based on the existing data at the beginning. Of course, such as the real-time emissions of some factory vehicles, building construction, etc. will also affect the ground PM2.5 concentration.
- However, due to the limited that we get information in the real world, we only estimate the ground $PM_{2.5}$ concentration on the basis of available data.
- Sampling selection, temporal variation, and spatial distribution based cross-validation demonstrated that the method presented here is skilled in providing reliable ground-level $PM_{2.5}$ concentrations with high spatial resolution (0.01°) and 24-hour temporal coverage, which is challenging especially for heavily polluted regions. Independent validations were also conducted for cloudy conditions and nightime, and no degradation of performance was found.
- We examined the importance of satellite observed AOD in the prediction of $PM_{2.5}$ during both daytime and nighttime using four sensitivity measures(Cortez and Embrechts, 2013), namely range S_r, gradient S_g, variance Sv, and average absolute deviation from the median S_{AAD}. We use sensitivity analysis and visualization to open the black box model of neural networks (Cortez and Embrechts, 2013). We evaluated the impact of each input parameter on the results. AOD accounts for more than 30% of the weight throughout the day, and the relative significance exhibits slightly higher values during nighttime (Table 5), emphasizing the importance of AOD observations in nighttime predictions. Land cover type and meteorological variables also play important roles in the dynamic evolution of $PM_{2.5}$ in North China and other regions, as illustrated in Figure. S14. The effects of the key variables on surface $PM_{2.5}$ concentrations are given in Figure. S15. However, the model tends to better capture moderately polluted conditions, as the relative errors exhibit relatively larger values when observed $PM_{2.5}$ concentrations are above 350 μg m^{-3} or below 20 μg m^{-3} (Figure. S16). The relatively poor capability of our ST-NN model in capturing these extremely low or high values are mainly attributed to the rarity of these conditions and the small sampling size (North China: 0.34‰, East China: 0.059‰, South China: 0.068‰, Sichuan Basin: 0.048‰, Shaanxi Province: 0.25‰). This is also the limitation of the data model. The lack of sample size in a specific case leads to a certain

deviation in the extreme case estimation of the model space to the real world space. Similar uncertainties of the model might be raised by the errors in model input data. Random errors were added to the input data to explore how it would influence the errors of predicted $PM_{2.5}$. Similarly, the quality of AOD data was essential within a very broad range of uncertainty (Figure. S17). When errors of other inputs grow (>20%), the accuracy of prediction would also be significantly degraded (Figure. S17).

- We examined also how input data quality control process would affect the accuracy, and a negligible role was found (Table S10, S11). During the development of ST-NN models for different regions in China, the loss function decreased in a similar manner, while the decreasing speed and convergence values varied among regions due to differences in the size and feature of data (Figure. S18). We noticed also that the performance of the model varied across regions and seasons, which might be also related to the distinct spatiotemporal features of $PM_{2.5}$ (Figure. S19), and the associated meteorological/geographical characteristics in different regions. The uneven distribution of CNEMC sites might also play a role (Figure. S13).
- A long-standing restriction for the use of satellite AOD has been that surface $PM_{2.5}$ cannot be constrained under cloudy conditions, during nighttime or during severe haze(Gao et al., 2017). This limitation has been overcome here with an advanced statistical method. The capability of the built ST-NN model in predicting $PM_{2.5}$ below clouds and during nighttime is mainly due to the consideration of spatiotemporal variation of influencing meteorological/ geographical factors and the dynamic evolution of aerosols. The processes considered are close to those in numerical chemical transport models, but with constraints of satellite AOD. Time-varying and time-invariant factors were processed separately in the ST-NN model to explore the dynamic feature of aerosol under complex influences, and the factors on different time scales were considered. Our ST-NN model relies on the regional transport features of air pollution, and it could thus be problematic to track very small point sources. This limitation will be further improved in future studies with more in-depth exploration of the connection between aerosol and clouds(Saide et al., 2012). The issue of rarely observed extreme conditions and small sampling size could be solved also to some extent in the near future when the volume of observations grows with time.

- Conclusion

- We built a ST-NN model by using multi-source heterogeneous data such as meteorology, elevation, population, etc., and obtained the ground $PM_{2.5}$ concentration results with full coverage of time and space (24-hour, hourly time resolution, 0.01° spatial resolution). It effectively fills the space and time gap of satellite and ground station monitoring. Through cross validation of samples, time and space, we get the result that the overall R^2 is greater than 0.8, and the slope of the estimate is close to 1, and the root mean square error is less than 26 μgm^{-3}.

And we open the black box model of neural network through sensitivity analysis and visualization methods, and quantitatively analyze the impact of each input variable on the results.”