

1 Spatiotemporal continuous estimates of daily 1-km PM_{2.5} from 2000 2 to present under the Tracking Air Pollution in China (TAP) 3 framework

4 Qingyang Xiao¹, Guannan Geng^{1*}, Shigan Liu², Jiajun Liu¹, Xia Meng³, Qiang Zhang²

5 ¹State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University,
6 Beijing 100084, China

7 ²Ministry of Education Key Laboratory for Earth System Modelling, Department of Earth System Science, Tsinghua
8 University, Beijing 100084, China

9 ³School of Public Health, Key Laboratory of Public Health Safety of the Ministry of Education and Key Laboratory of Health
10 Technology Assessment of the Ministry of Health, Fudan University, Shanghai 200032, China

11 *Correspondence to:* Guannan Geng (guannangeng@tsinghua.edu.cn)

12 **Abstract.** High spatial resolution PM_{2.5} data covering a long time period are urgently needed to support population exposure
13 assessment and refined air quality management. In this study, we provided complete-coverage PM_{2.5} predictions with a 1-km
14 spatial resolution from 2000 to the present under the Tracking Air Pollution in China (TAP, <http://tapdata.org.cn/>) framework.
15 To support high spatial resolution modelling, we collected PM_{2.5} measurements from both national and local monitoring
16 stations. To correctly reflect the temporal variations in land cover characteristics that affected the local variations in PM_{2.5}, we
17 constructed continuous annual geoinformation datasets, including the road maps and ensemble gridded population maps, in
18 China from 2000 to 2021. We also examined various model structures and predictor combinations to balance the computational
19 cost and model performance. The final model fused 10-km TAP PM_{2.5} predictions from our previous work, 1-km satellite
20 aerosol optical depth retrievals, and land use parameters with a random forest model. Our annual model had an out-of-bag R²
21 ranging between 0.80 and 0.84, and our hindcast model had a by-year cross-validation R² of 0.76. This open-access 1-km
22 resolution PM_{2.5} data product with complete coverage successfully revealed the local-scale spatial variations in PM_{2.5} and could
23 benefit environmental studies and policy-making.

26 **1 Introduction**

27 Air pollution has been a nonnegligible environmental problem around the world. China implemented strict clean air policies
28 in the past decade that considerably improved air quality. To support the policy evaluation and air quality management, we
29 constructed the Tracking Air Pollution in China (TAP) platform (<http://tapdata.org.cn/>), which provides near real-time air
30 pollutants, i.e., PM_{2.5} and O₃, distribution at a 0.1 degree (approximately 10 km) spatial resolution, from the fusion of ground
31 measurements, satellite retrievals, and chemical transport model (CTM) simulations (Geng et al., 2021). The TAP data
32 benefited the evaluations of clean air policies and the characterization of air pollution exposure (Xiao et al., 2020; Xiao et al.,
33 2021b; Xiao et al., 2021c). However, with the improved air pollution control targets that require refined air quality management,
34 the detailed monitoring of air pollution distribution at higher spatial resolutions is urgently needed.

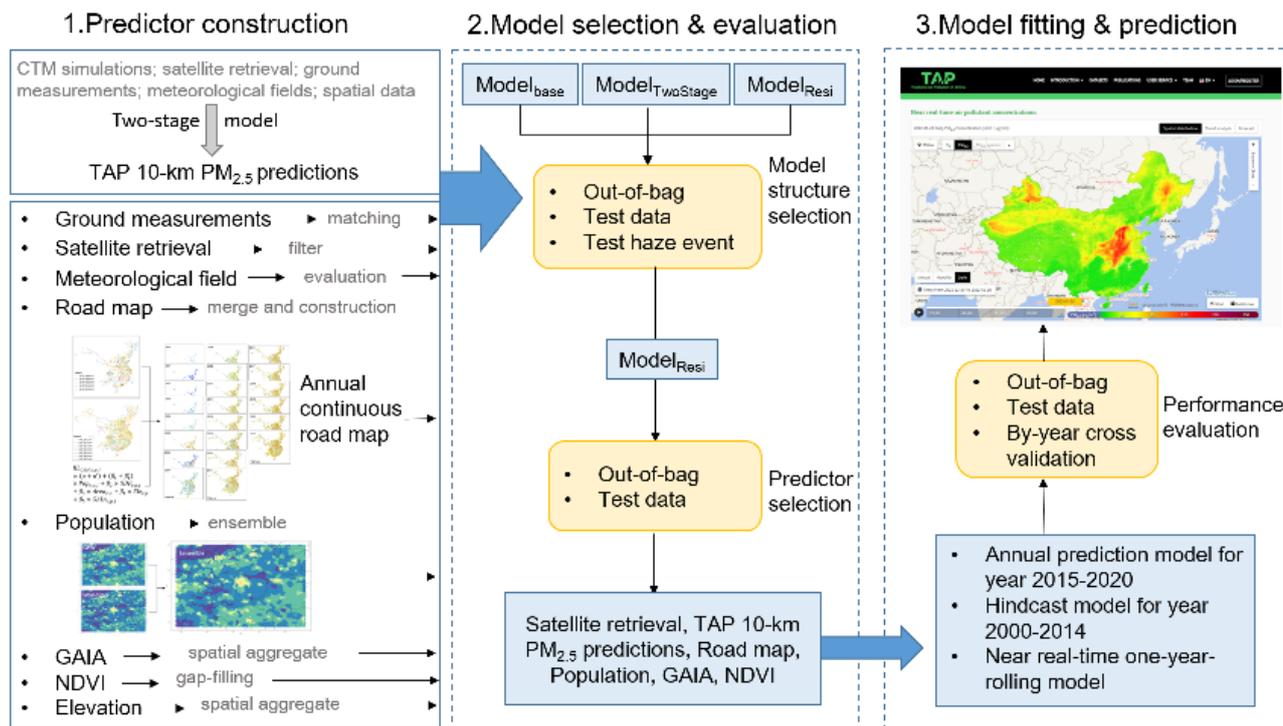
35 Recent developments in machine learning algorithms and remote sensing techniques have fueled the production of air pollution
36 data at high spatiotemporal resolutions. For example, moderate-resolution imaging spectroradiometer (MODIS) products
37 provide aerosol optical depth (AOD) retrievals at a 3-km resolution, contributing to the prediction of ground-level PM_{2.5}
38 concentrations at the local scale (Xie et al., 2015; He and Huang, 2018; Hu et al., 2019). The multi-angle implementation of
39 atmospheric correction (MAIAC) algorithm provides AOD retrievals at a 1-km resolution and benefits predictions of PM_{2.5}
40 distribution at a 1-km (Wei et al., 2021; Goldberg et al., 2019; Xiao et al., 2017; Bai et al., 2022b) or higher spatial resolution
41 (Huang et al., 2021). Recently, with the Gaofen-5 satellite retrieval, Zhang et al. (2018) predicted the PM_{2.5} concentration at a
42 160-m resolution. However, most of these high-resolution data products are limited to after 2013 or cover a specific region of
43 China. Few studies have filled the missing predictions resulted from missing satellite retrievals (Bai et al., 2022b; Ma et al.,
44 2022). This discontinuous PM_{2.5} prediction in space and time not only limits the application of PM_{2.5} products in scientific
45 research and environmental management but also biases the characterization of population exposure to PM_{2.5} pollution (Xiao
46 et al., 2017). Additionally, although high-resolution PM_{2.5} prediction models widely included various land use data, e.g., road
47 maps, land cover types, and points of interest (POIs), to describe the local-scale spatial variations in air pollution emissions
48 and air pollution levels, most studies used only one or two years of land use data during the whole study period and ignored
49 the critical variations in them. This lack of temporal variations in land use data may affect the spatial accuracy of high-
50 resolution PM_{2.5} predictions.

51 In this study, we constructed a high-resolution PM_{2.5} concentration prediction system under the TAP framework in order to
52 provide 1-km resolution full-coverage PM_{2.5} retrievals covering a long time period. To correctly reveal the spatial
53 characteristics of PM_{2.5} distribution at such a high spatial resolution, we processed MAIAC satellite retrievals as well as
54 evaluated and constructed various temporally continuous land-use parameters with statistical and geospatial modelling that
55 have not been included in previous TAP models. By fusing high-resolution MAIAC satellite retrievals, TAP PM_{2.5} products at
56 a 10-km resolution, satellite normalized difference vegetation index (NDVI) products, and various continuous long-term land

57 use data, we provide 1-km resolution $PM_{2.5}$ predictions from 2000 to the present with complete coverage and timely updates.
 58 The high quality and easy accessibility of our high-resolution $PM_{2.5}$ data could support research on air pollution and
 59 environmental health at local scales and contribute to the management of local air quality.

60 2 Data and Method

61 The workflow of this study is shown in Fig. 1. First, we processed and assimilated various predictors. The daily scale varying
 62 predictors include satellite retrieval, TAP 10-km $PM_{2.5}$ predictions, and meteorological fields, and the land use variables
 63 include road map, population distribution, artificial impervious area, and vegetation index. In China, the high-speed economic
 64 development in the past several decades has led to significant changes in land use and population distribution. To correctly
 65 reveal the temporal variations in land use parameters and further benefit the description of local-scale $PM_{2.5}$ concentration
 66 variations, we constructed temporally continuous land use predictors through statistical and spatial modeling. We then
 67 optimized the model structure and selected model predictors through various examinations to balance the model performance
 68 and computing cost. With the selected model design, we fitted three models under the TAP framework: the hindcast model
 69 with training data from 2013 to 2020 to predict historical $PM_{2.5}$ concentrations from 2000 to 2014; the annual model with
 70 training data of each corresponding year from 2015 to 2020; and the near real-time model with rolling one-year training data
 71 that provides near real-time $PM_{2.5}$ predictions until the day before present day.



72

73 **Figure 1 Workflow of this study.**

74 **2.1 Ground measurements of PM_{2.5}**

75 The hourly PM_{2.5} measurements from national air quality monitoring stations were downloaded from the China National
76 Environmental Monitoring Center (<http://www.cnemc.cn/>). To examine the model's prediction ability of in space, we also
77 collected PM_{2.5} measurements from local air quality monitoring stations operated by local government agencies
78 (Supplementary Fig. 1). The hourly PM_{2.5} concentration measurements below 1 µg/m³, the lowest measurement limit of most
79 monitors, and above 2000 µg/m³ were excluded for quality control of measurements. Identical continuous measurements
80 during at least three hours were also removed. Hourly data generated daily average PM_{2.5} concentrations with fewer than 18
81 hourly measurements were removed.

82 In order to examine whether quality of measurements from national monitors and from regional monitors is comparable, we
83 matched the nearest national and regional monitors to compare their measurements. Specifically, we selected such national-
84 regional monitor pairs with distance less than 0.5 degree between them and compared their daily average PM_{2.5} measurements.
85 This comparison illustrates that measurements from regional monitors matched well with measurements from the nearest
86 national monitors with the linear regression R² of 0.89 (slope of 0.99 and intercept of 1.00). The average difference between
87 daily matched regional and national measurements are 1.6 µg/m³.

88 **2.2 Full-coverage PM_{2.5} predictions at a 10-km resolution**

89 The 10-km resolution PM_{2.5} predictions, which were estimated from a two-stage machine learning modeling system, were
90 downloaded from the TAP website (<http://tapdata.org.cn/>) (Xiao et al., 2021c). Previous evaluations reported that the PM_{2.5}
91 prediction model with the out-of-bag R² (the R² of the linear regression between measurements and predictions from trees that
92 did not include these measurements for training) ranged between 0.80 and 0.88 (Geng et al., 2021). These TAP PM_{2.5} data at
93 a 10-km resolution with complete coverage are updated in near real-time.

94 **2.3 MAIAC retrievals**

95 The multi-angle implementation of atmospheric correction (MAIAC) (Lyapustin et al., 2011a; Lyapustin et al., 2011b) data
96 were downloaded from the NASA Earthdata site (<https://lpdaac.usgs.gov/products/mcd19a2v006/>). Only pixels with high-
97 quality retrievals were included (QA. CloudMask=clear and QA. AdjacencyMask=clear) (Kloog et al., 2015; Lyapustin, 2018).
98 Since the Aqua and Terra satellites crossover at around 10:30 am and 1:30 pm local time respectively, the daily
99 spatial missing patterns of Aqua and Terra AOD retrievals are different. To improve the coverage of AOD, we
100 fitted daily linear regressions between Aqua AOD and Terra AOD. Then we predicted the missing AOD of one
101 satellites when the AOD of another satellite exists. After the daily linear interpolation, the Aqua and Terra AODs were
102 averaged to reflect the daily aerosol loadings.

103 2.4 Meteorological fields and evaluation

104 Various reanalysis data products, including ERA5, MERRA-2, and GEOS-FP, have been used to provide meteorological files
105 in previous air pollution prediction models (Geng et al., 2021;Wei et al., 2021;Xiao et al., 2017). In this study, to select the
106 best-performing meteorological data with long temporal data coverage (from 2000 to the present) and timely updating, we
107 evaluated ERA5-Land, ERA5, and MERRA-2 meteorological datasets with meteorological measurements at regional air
108 quality monitoring stations during 2019. The evaluation results showed that ERA5-Land data at a 0.1-degree resolution
109 outperformed the MERRA-2 reanalysis data and the ERA5 reanalysis data (Supplementary Table 1). We extracted and
110 processed the ERA5-Land parameters, including 2-m temperature, 10-m u- and v-component of wind, surface pressure, and
111 total precipitation, for model predictor selection. The 2-m relative humidity was calculated from the 2-m temperature and the
112 2-m dew point temperature.

113 2.5 Construction of the time series of land use variables

114 2.5.1 Population density

115 We evaluated and fused various global gridded population density data that were publicly available, including the LandScan
116 Global Population Database from 2000 to 2019 (Dobson et al., 2000); the Gridded Population of the World (GPW) data product
117 (version 4) for 2000, 2005, 2010, 2015, and 2020 (Doxsey-Whitfield et al., 2015); and the annual WorldPop data at a 1-km
118 resolution from 2000-2020 (Wardrop et al., 2018;Reed et al., 2018). We linearly interpolated the GPW data for each year. For
119 data quality evaluation, we obtained the sum population at the county or city level from 2000 to 2019 from the China City
120 Yearbooks. The gridded population datasets were aggregated to county- or city sums and compared to the yearbook records.
121 The LandScan data outperformed the other two datasets (Supplementary Fig. 2); however, the spatial distribution of the
122 LandScan data showed an unreasonable pattern and were excluded. As shown in Supplementary Fig. 4, the LandScan
123 data present very high population density along the road and many randomly-distributed high-population grids in
124 certain square areas on the map. Due to the significant spatial variations in data accuracy (Wang et al., 2004;Bai et al.,
125 2018), we fused the GPW and WorldPop data to improve data quality across space. We first fitted linear regressions between
126 gridded population and yearbook records of each county or city that has at least 6 matched data pairs. Then, we averaged the
127 gridded population with the coefficient of determination (R^2) as a weight (Eq. (1)). We selected the R^2 rather than the root
128 mean square error as the weight because the spatial variation trends were more important than the number of populations in
129 the prediction of $PM_{2.5}$.

$$130 \text{Pop}_{g(i),y} = \frac{GPW_{g(i),y} \times R_{GPW,i}^2 + WorldPop_{g(i),y} \times R_{WorldPop,i}^2}{R_{GPW,i}^2 + R_{WorldPop,i}^2} \quad (1)$$

131 where $Pop_{i,y}$ represents the ensemble population of grid g in county/city i of year y ; $GPW_{g(i),y}$ and $WorldPop_{g(i),y}$ represent
132 the population of grid g year y of dataset GPW and WorldPop, respectively; and $R_{GPW,i}^2$ and $R_{WorldPop,i}^2$ represent R^2 of GPW
133 and WorldPop in county/city i .

134 For counties/cities that did not have sufficient matched data pairs for regression fitting, we employed the weight of the nearest
 135 county/city for the ensemble (Supplementary Fig. 3). We subsequently constrained the city level and national sum population
 136 to be consistent with the record from the China City Statistical Yearbook and the China Statistical Yearbook.

137 **2.5.2 Land cover**

138 The percentage of artificial impervious area of each 1-km modeling grid was calculated from the annual global artificial
 139 impervious area (GAIA) data at a 30-m resolution from 2000 to 2018 (<http://data.ess.tsinghua.edu.cn/gaia.html>). To
 140 estimate the GAIA distribution after 2018, we fitted linear regressions with data from 2016 to 2018 for each grid and
 141 extrapolated the GAIA values of 2019, 2020, and 2021. The data from 2013 to 2017 were used to evaluate the performance of
 142 this linear extrapolation. The examination results comparing the GAIA data and the first-year, second-year, and third-year
 143 extrapolated GAIA predictions showed that the R^2 values ranged from 0.996–0.999, 0.985–0.989, and 0.969–0.979,
 144 respectively.

145 **2.5.3 Road map**

146 Limited road maps are available in China. We collected the annual road maps from 2013 to the present from OpenStreetMap
 147 (www.openstreetmap.org) (Barrington-Leigh and Millard-Ball, 2017), a crowdsourced collaborative geographic information
 148 collection project. OpenStreetMap data have been widely used for road density analysis and the construction of world road
 149 data products (Zhang et al., 2015; Meijer et al., 2018). A previous evaluation study reported on the considerable accuracy of
 150 the OpenStreetMap data (Haklay, 2010). To evaluate the quality of the OpenStreetMap data and to fill the historical data gap
 151 before 2013, we also collected road maps of 2000, 2004, 2005, 2010, 2012, 2014, and 2015 from the survey.

152 We first extracted the length of various types of roads from the OpenStreetMap data and from the road survey at the grid level.
 153 We combined some road types to make the road classification more comparable in OpenStreetMap and in the survey
 154 (Supplementary Table 2). To estimate the annual road distribution before 2013, we first compared the grid-level road length
 155 of 2014 extracted from OpenStreetMap and the survey map (Supplementary Table 3). Then, we modify the survey map data
 156 to construct the OpenStreetMap-type gridded road length of the years that the survey map data were available by the equation
 157 listed below:

$$158 \quad RL_{OSM,i,j} = \frac{RL_{OSM,i,2014} \times RL_{road,i,j}}{RL_{road,i,2014}} \quad (2)$$

159 where $RL_{OSM,i,j}$ represents the OpenStreetMap-type road length of year j over grid i and $RL_{road,i,j}$ represents the survey-type
 160 road length of year j over grid i .

161 After estimating the OpenStreetMap-type road length of years when the survey maps were available, we filled the gap years
 162 by weighted linear interpolation. First, we estimated the city-level sum road length of different road types by a linear mixed
 163 effects model (LME)(Meijer et al., 2018):

$$164 \quad RL_{OSM,c,p,j} = (\mu + \mu') + (\beta_1 + \beta_1') \times Pop_{c,p,j} + \beta_2 \times GDP_{c,p,j} + \beta_3 \times Area_{c,p} + \beta_4 \times Ele_{c,p} + \beta_5 \times GAIA_{c,p,j} \quad (3)$$

165 where $RL_{OSM,c,p,j}$ represents the sum road length of city c in province p of year j ; μ represents the fixed intercept and μ'
166 represents the province-level random intercept; $\beta_1, \beta_2, \beta_3, \beta_4$, and β_5 represent the fixed slope of the city average population
167 density, per capita gross domestic product (GDP), city area, city average elevation and city average GAIA; β'_1 represents the
168 province-level random slope of population density. The \log_{10} transformation was conducted for all the continuous variables to
169 account for the skewed distribution (Meijer et al., 2018). Stepwise linear regression was used to select the significant predictors
170 (Meijer et al., 2018). To evaluate the LME model performance, by-year cross validation and four-year cross validation were
171 conducted. Regarding the by-year cross-validation, we selected one-year data for model testing in sequence and used the data
172 of the remaining years for model training. Regarding the four-year cross validation, we selected 2013, 2014, and 2015 for
173 model testing in sequence and used the data four years after the corresponding testing year for model fitting. For example,
174 when selecting 2013 for model testing, the data from 2017, 2018, and 2019 were used to fit the model.
175 After estimating the city-level sum road length, we further used the sum road length as a weight to assign the road length
176 changes to each gap year. The equation is listed below:

$$177 \quad RL_{OSM,i,j} = RL_{OSM,i,start} + (RL_{OSM,i,end} - RL_{OSM,i,start}) \times \frac{RL_{OSM,c,j} - RL_{OSM,c,start}}{RL_{OSM,c,end} - RL_{OSM,c,start}} \quad (4)$$

178 where $RL_{OSM,i,start}$ and $RL_{OSM,i,end}$ represent the road length over grid i of the starting and ending year with available OSM-
179 type road data, respectively; $RL_{OSM,c,start}$ and $RL_{OSM,c,end}$ represent the road length of city c of the starting and ending year
180 with available OSM-type road data, respectively; and $RL_{OSM,c,j}$ represent the road length of city c of the gap year j .

181 2.6 Other auxiliary datasets

182 We downloaded the monthly Terra MODIS NDVI (MOD13A3) at a 1-km resolution and filled the missing NDVI data by the
183 nearest neighbour spatial smoothing approach. The average elevation data at a 30-m resolution were extracted from the
184 Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM)
185 version 2.

186 2.7 Data assimilation

187 All the predictors were assimilated to the 1-km MAIAC pixels by various geostatistical methods. The meteorological data and
188 $PM_{2.5}$ predictions at a 0.1 degree resolution were downscaled to the 1-km MAIAC pixels with the inverse distance weighting
189 method. The elevation pixels falling in each 1-km grid cell were averaged. The NDVI data were assigned to the MAIAC pixels
190 by the nearest neighbour method.

191 2.8 Optimization and evaluation of the prediction model

192 To make the $PM_{2.5}$ prediction process efficient and highly accurate, we designed various examinations to optimize the model
193 structure and identify key predictors of the $PM_{2.5}$ prediction system.

194 Three model structures were evaluated: $model_{TwoStage}$ has a two-stage design with the first stage model predicting the high
195 pollution indicator and the second stage model predicting the residual between 10-km TAP $PM_{2.5}$ predictions and
196 measurements (Xiao et al., 2021c); $model_{Resi}$ is a one-stage model that predicts the residual between 10-km TAP $PM_{2.5}$
197 predictions and measurements; and $model_{Base}$ is a one-stage model that directly predicts the $PM_{2.5}$ concentration with the 10-
198 km TAP $PM_{2.5}$ prediction as a predictor. Since the underestimation of high pollution events are widely reported, in addition to
199 the evaluations including all the test measurements, we conducted additional evaluations focusing on the prediction accuracy
200 of haze events when the daily average $PM_{2.5}$ concentration was higher than the $75 \mu\text{g}/\text{m}^3$ national secondary air quality standard.
201 Then, we selected the critical predictors of the $PM_{2.5}$ prediction model (Fig. 1). We first constructed the full model with all the
202 predictors and then we removed the meteorological predictors in sequence, according to the importance of parameters
203 estimated from the full model. Predictors with the smallest importance were removed first. Data from 2019 were used for
204 model predictor optimization.

205 Various statistics were employed to characterize model performance. The out-of-bag (OOB) predictions were provided during
206 the training of the random forest that the measurements were predicted by trees that were trained with randomly selected
207 samples without each of these measurements. Comparing the out-of-bag predictions with measurements in linear regression
208 provided us with the out-of-bag R^2 , the root mean square error (RMSE), and the mean prediction error (MPE). To evaluate the
209 model's ability to reveal $PM_{2.5}$ variations at the local scale and at locations without monitoring, we used the measurements
210 from national stations for model training and the measurements from the high-density local stations from model evaluation.
211 These local stations are primarily located in Hebei, Henan, Shandong, and Chengdu (Supplementary Fig. 1). The evaluation
212 results of OOB and with test data were used to optimize the model structure and select predictors. Then, to evaluate the
213 optimized final model's prediction ability in time, we conducted a by-year cross validation.

214 Consistent with the previous TAP $PM_{2.5}$ prediction framework, the missing satellite AOD retrievals were filled by adjusting
215 the first layer of the decision tree and setting the availability status of AOD as the cutoff point of the first layer of the decision
216 trees. The performance of this gap-filling method has been fully evaluated in our previous studies (Xiao et al., 2021a; Geng et
217 al., 2021)

218 **3 Results**

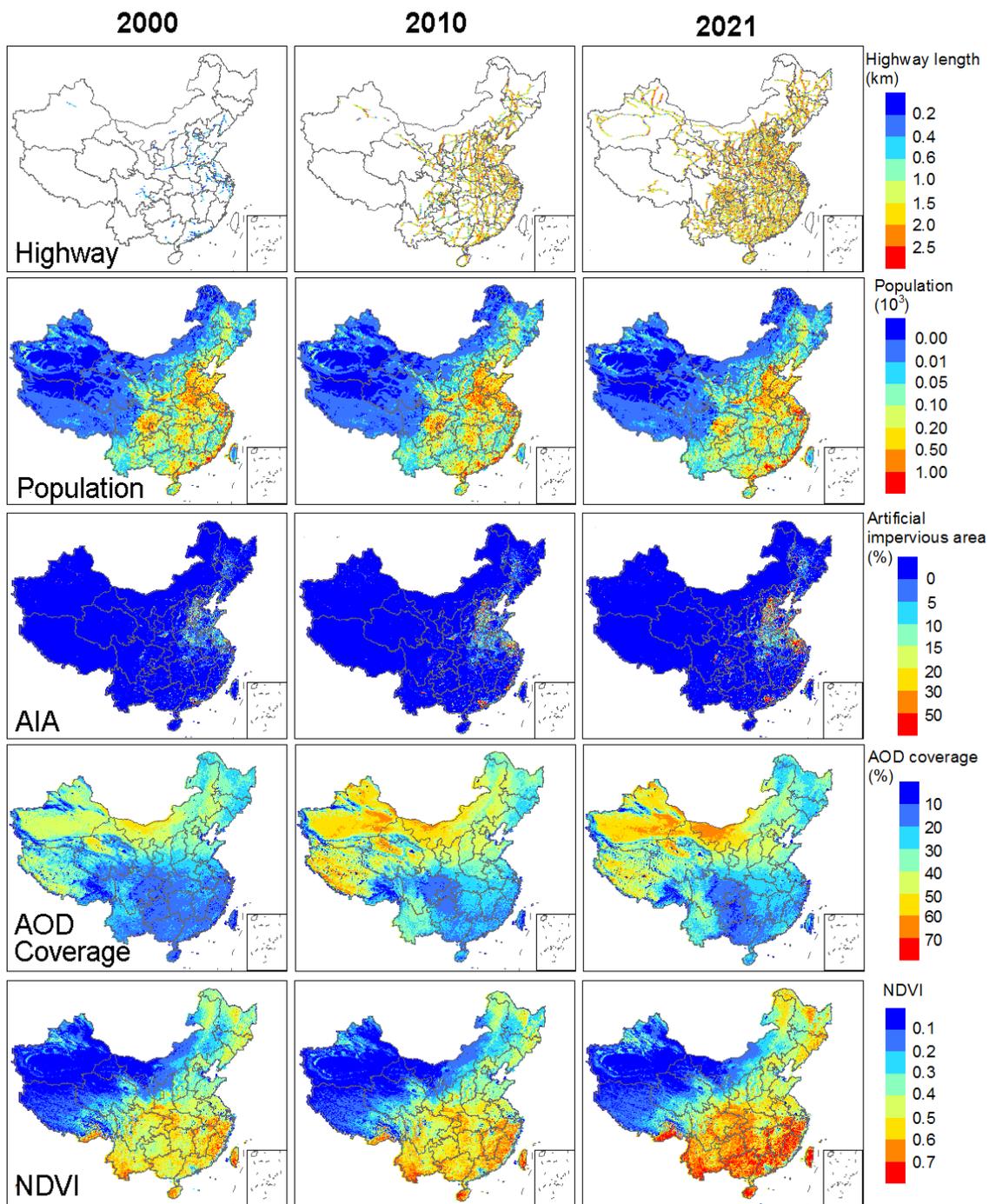
219 **3.1 Temporal variations in predictors**

220 The high-resolution $PM_{2.5}$ prediction was supported by various high-resolution predictors. In addition to the 1-km resolution
221 MAIAC AOD, we also constructed and presented various 1-km resolution predictors, including road map, population
222 distribution, artificial impervious area, and NDVI (Fig. 2).

223 We evaluated the road length model for various road types through by-year cross validation and four-year cross validation
224 (Supplementary Table 4). The cross-validation predictions of all road types were highly correlated with the OSM data. The
225 four-year cross-validation performance were comparable to the by-year cross-validation performance, indicating that the

226 model's temporal prediction ability was robust. The performance of the secondary road model was slightly better than that of
227 the highway model and primary road model, showing higher correlations between local socioeconomic factors and secondary
228 road length relative to highways and primary roads that are constructed nationally. The predicted highway length correctly
229 revealed the temporal trends of the records of highway length from the China traffic yearbooks, with correlation coefficients
230 of 0.99. Since the road type classifications of the OSM and China traffic yearbooks are inconsistent, we did not compare the
231 lengths of other types of roads. We observed a consistent increasing trend in road length for all road types across China (Fig.
232 2). The predicted road maps also displayed the construction of some local landmarks, e.g., the 6th Ring Road in Beijing.
233 Compared to the statistical yearbook records, our ensemble population data showed R^2 and root mean square error (RMSE)
234 values of 0.74 and 0.19 million, respectively, outperforming other gridded population data (Supplementary Fig. 5). The
235 changes in population density distribution were inconsistent across China due to the substantial internal migration during the
236 past decades (Fig. 2). For example, we observed that the high-speed economic development in Shenzhen city and the whole
237 Pearl River Delta (PRD) attracted a large migration population, but the populations in small cities in Northeast and Central
238 China were consistent or decrease. The artificial impervious area also increased significantly across China, especially in
239 regions with fast economic development. The consistent increase in NDVI over most parts of China, especially in the southeast,
240 showed the achievement of environmental protection in China. We found considerable missingness in MAIAC retrievals over
241 China, especially in the southeast and northeast regions with large populations. Thus, gap-filling is necessary to provide
242 valuable $PM_{2.5}$ predictions across China.

243



244

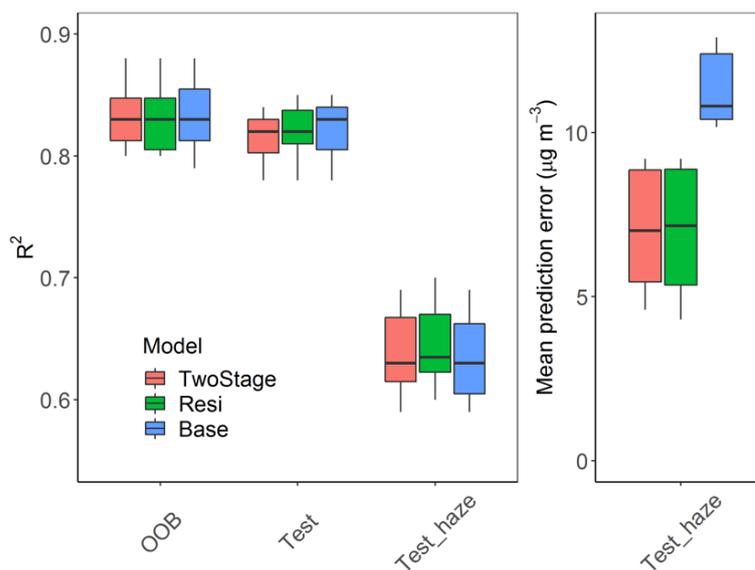
245

246

Figure 2 Estimated annual distributions of key predictors, including highway length, population, artificial impervious area, MAIAC AOD coverage, and normalized difference vegetation index (NDVI), in 2000, 2010, and 2021 across China.

248 3.2 Optimization of the high-resolution PM_{2.5} prediction model

249 Three model structures, model_{TwoStage}, model_{Resi}, and model_{Base}, were examined in this study. The evaluation results showed
 250 that these candidate models performed comparably in R² in all the evaluations (Fig. 3). However, the model_{Base} that directly
 251 predicts the measurements showed significantly larger prediction error than the other two models during haze events. For some
 252 years, e.g., 2017 and 2018, the average prediction error of the model_{Base} was more than double than the prediction error of
 253 model_{TwoStage} and model_{Resi}. This result was consistent with our previous findings that the prediction of residuals enlarges the
 254 response of the dependent variable to the predictors, thus benefiting the prediction of extreme events(Xiao et al., 2021c). We
 255 did not observe significant differences between model_{TwoStage} and model_{Resi}. Thus, considering the prediction performance and
 256 the model fitting time expense, the model_{Resi} was selected.



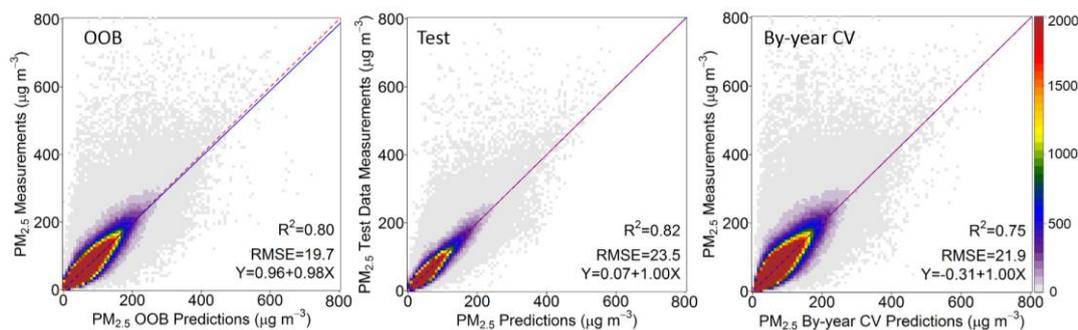
257

258 **Figure 3** The performance of model_{TwoStage}, model_{Resi}, and model_{Base} in the out-of-bag evaluation (OOB), the evaluation with test data
 259 from local stations (Test) and the evaluation with test data higher than 75 µg/m³ (Test_haze). Left: the evaluation R² Right: the
 260 average prediction error.

261 We then examined the contribution of meteorological fields to the high-resolution PM_{2.5} prediction (Supplementary Table 5).
 262 Compared to the full model, the out-of-bag R² of the model without meteorological fields decreased from 0.85 to 0.80; however,
 263 the R² with test data decreased only 0.02, from 0.85 to 0.83. This evaluation showed that the contribution of meteorological
 264 fields to high-resolution PM_{2.5} predictions was limited. Potential reasons include the coarse resolution of the meteorological
 265 data limiting the characterization of high-resolution variations in meteorological fields or in PM_{2.5} distributions. Additionally,
 266 the meteorological effects on PM_{2.5} have been considered in the 0.1-degree PM_{2.5} data that served as a predictor in the model.

267 Comprehensively considering the model performance and the meteorological data update frequency, we removed
 268 meteorological fields.

269 Table 1 summarizes the out-of-bag performance of our final annual models and hindcast model. The model R^2 ranged between
 270 0.80 and 0.84 for annual models. The small interannual variations indicated that our model was robust and provided predictions
 271 with constant quality during the study period. The very small model mean prediction error (bias) together with the slopes close
 272 to 1 showed the inexistence of systemic bias in the prediction models. Our model performance was comparable with previous
 273 studies (Huang et al., 2021; Wei et al., 2019; Wei et al., 2020; Wei et al., 2021). The highcast model performed comparably in
 274 the OOB evaluation, the test data evaluation, and the by-year cross-validation evaluation (Fig. 4), showing great accuracy and
 275 high robustness. No considerable overfitting was observed, and no system bias was detected in the spatial prediction (test data)
 276 and temporal prediction (by-year cross-validation) examinations.



277
 278 **Figure 4 Model performance of the hindcast model trained with all the data from 2013 to 2019**

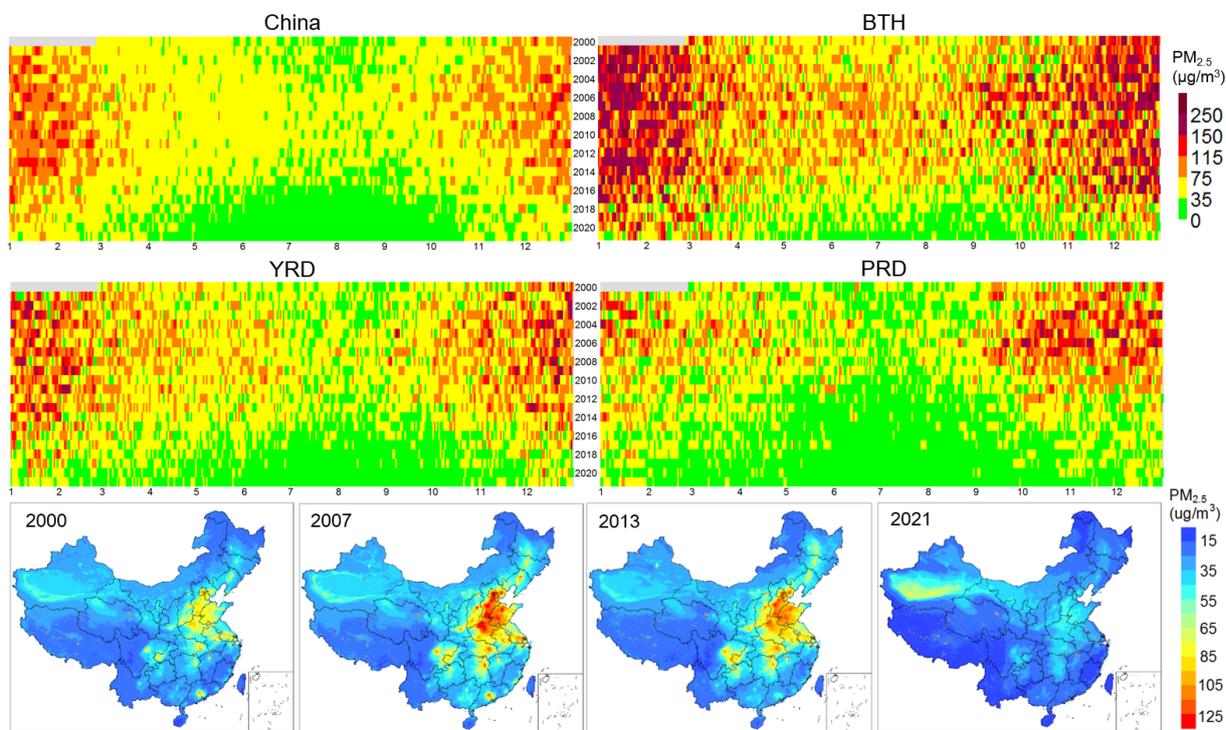
279 **Table 1 Out-of-bag performance of the annual model trained with all data of each year and the hindcast model trained with all data**
 280 **during 2013-2019**

| Model | R^2 | slope | RMSE (μg^3) | Bias (μg^3) |
|-------------|-------|-------|--------------------------|--------------------------|
| Annual-2015 | 0.82 | 0.99 | 20.20 | -0.06 |
| Annual-2016 | 0.83 | 1.01 | 18.24 | -0.05 |
| Annual-2017 | 0.84 | 1.00 | 16.67 | -0.02 |
| Annual-2018 | 0.82 | 0.95 | 15.94 | -0.01 |
| Annual-2019 | 0.81 | 0.95 | 16.35 | -0.04 |
| Annual-2020 | 0.80 | 0.95 | 14.96 | -0.02 |
| Hindcast | 0.80 | 0.98 | 19.6 | -0.03 |

281 **3.3 The spatiotemporal characteristics of the high-resolution PM_{2.5} map**

282 The high-resolution PM_{2.5} maps revealed critical local patterns of annual (Fig. 5-6) and daily (Fig. 7-8) pollution distributions
 283 that could not be identified by the 0.1-degree resolution maps. Comparing the daily population weighted average PM_{2.5}
 284 concentrations from 2000 to 2021, the number of days with PM_{2.5} higher than 75 $\mu\text{g}/\text{m}^3$ were significantly reduced after 2013

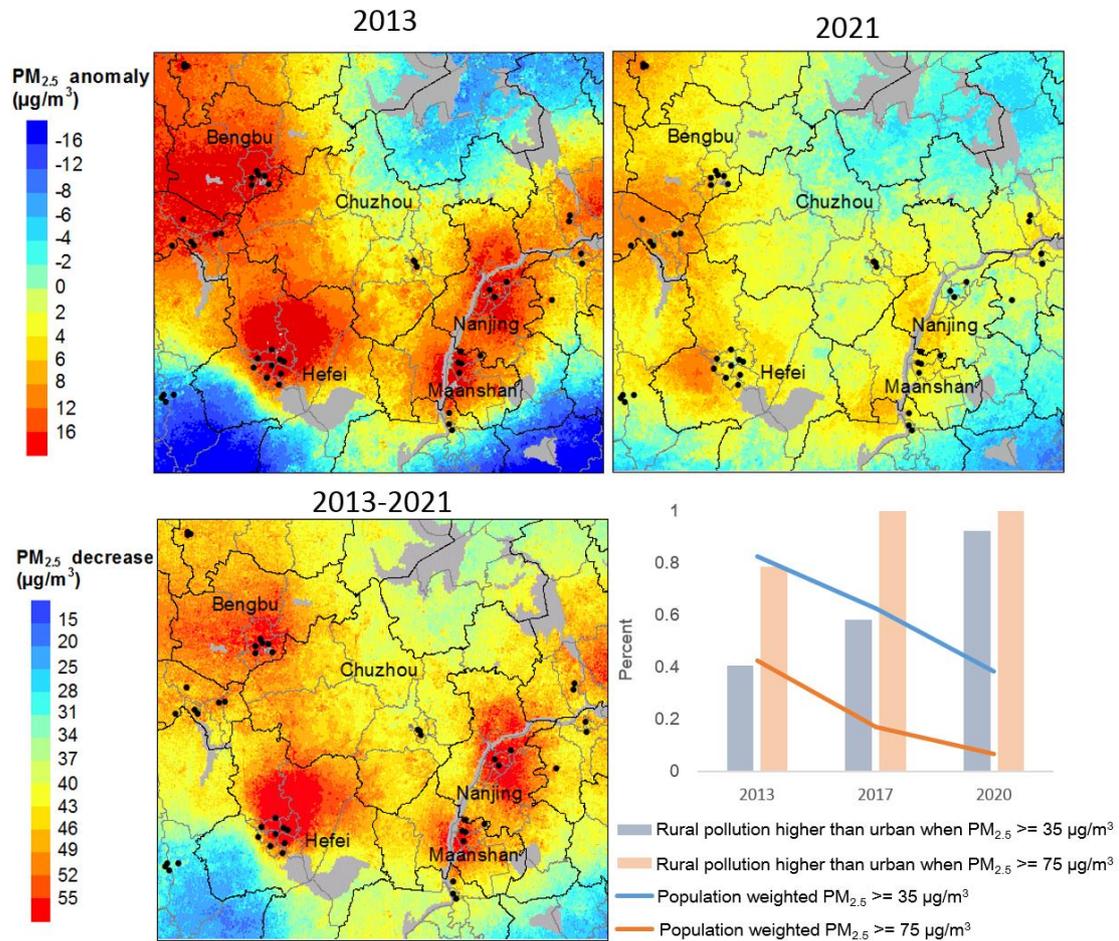
285 across China (Fig. 5). BTH showed high pollution levels with the haze days occurred across the year. In recent years, benefited
 286 from the pollution control policies, high pollution days in BTH outside winter were basically removed. The annual maps of
 287 $PM_{2.5}$ distribution in 2000 showed that the most polluted regions were located in Beijing, Hebei, and north of Henan; in 2007
 288 and 2013, the highly polluted regions extended and covered BTH, Shandong, Shanxi, Hunan, Sichuan basin, and YRD. After
 289 2013 with the strict pollution control policies, the air quality across China was significantly improved and the polluted regions
 290 shrunk in 2021.



291
 292 **Figure 5 Temporal variations of the daily population weighted $PM_{2.5}$ cover China, BTH, YRD, and PRD as well as the annual average**
 293 **$PM_{2.5}$ distribution in 2000, 2007, 2013, 2021.**

294 Figure 6 highlighted the variations in spatial distribution of $PM_{2.5}$ at the local scale. We compared the annual $PM_{2.5}$ anomaly,
 295 which is the gridded $PM_{2.5}$ minus the regional average $PM_{2.5}$, in 2013 and 2021 in YRD. The pollution hotspots transferred
 296 from city center with monitors to rural regions with limited monitoring. We found that after 2013, although the percent of days
 297 and counties with population weighted $PM_{2.5}$ violated than the primary ($35 \mu g/m^3$) and secondary air quality standard are
 298 continuously decreasing, the percent of days and counties with rural-county pollution higher than urban-county pollution
 299 significantly increased. In 2013, more than half of days and counties showed higher pollution in urban counties than in rural
 300 counties when the $PM_{2.5}$ was greater than $35 \mu g/m^3$; however, in 2021, more than 96% of this pollution days and counties
 301 showed lower pollution in urban counties than in rural counties. In 2017 and 2020, all the days with $PM_{2.5}$ greater than 75
 302 $\mu g/m^3$ showed higher rural-county pollution than urban-county pollution. One reason of the transportation of pollution hotspot

303 is that the $PM_{2.5}$ reduction during 2013-2021 was much greater in city centers than in rural regions. In 2021, most regional
 304 high pollution hotspots were transferred to around the junction of cities or towns.
 305



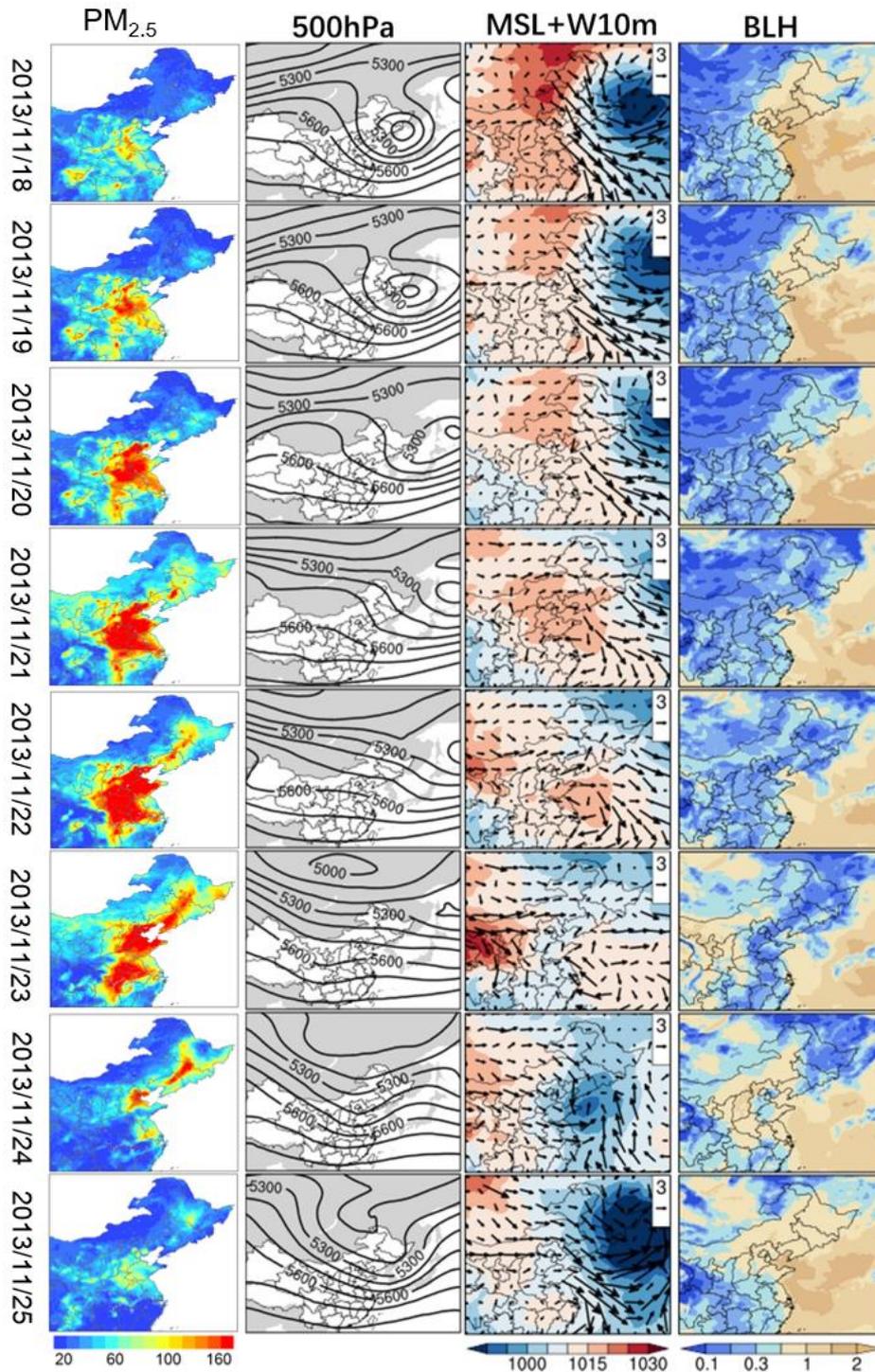
306

307 **Figure 6** The spatial distribution of annual average $PM_{2.5}$ anomaly in 2013 and 2021 in YRD (upper) as well as the changes in annual
 308 average $PM_{2.5}$ between 2013 and 2021 (lower left). The national air quality monitors are shown as dots; and the temporal trends in
 309 percent of days and counties with rural-county pollution higher than urban-county pollution in this region.

310 The daily maps showed more short-term local pollution variations. Figure 7 shows one haze event during November 18–25,
 311 2013. Since November 18th, the upper cyclone moved toward northwest and the North China Plain was covered by high-
 312 pressure ridge with continuously strengthen downdraft, leading to stable atmosphere that was unfavorable for pollution control.
 313 From November 18–23, the pollution kept increasing and triggered the haze event. Then, since November 24th, with the upper-
 314 level ridge moved eastward to the ocean, the North China Plain was affected by the trough with increased vertical upward

315 movement and raised boundary layer height. Both the horizontal and vertical diffusion was improved and the pollutant
316 concentrations decreased sharply, leading to the end of this haze event.

317 The 1-km resolution pollution map was able to reveal regional characteristics. For example, the impact of local transportation
318 emissions was observed on some days in the populous key regions (Fig. 8). These maps also indicated the importance of
319 including time-varying land use data for air pollution predictions, especially in high-resolution predictions, since the land use
320 characteristics led to noticeable spatial variations in the pollution distribution, as expected. To examine the impact of using
321 temporally mismatched land use data on the retrieved spatial patterns of $PM_{2.5}$, we used historical land use data (GAIA, NDVI,
322 population and road map of 2000) to predict the daily $PM_{2.5}$ distribution over these key regions of the same day. The historical
323 land use data in 2000 led to incorrect spatial characterizations of the $PM_{2.5}$ distribution in 2015.

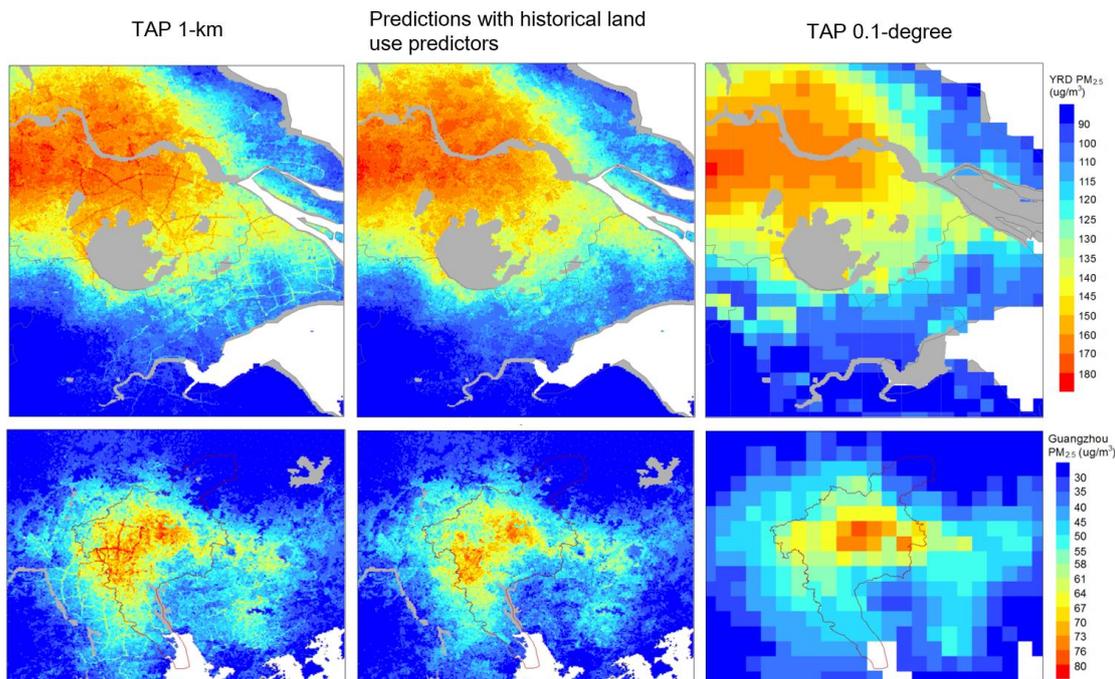


324

325

326

Figure 7 Daily $PM_{2.5}$ and meteorological field distributions during November 18–25, 2013. 500hPa: the vertical height at which the pressure is 500 hPa; MSL: mean sea level; W10m: wind speed and direction at 10 m; BLH: boundary layer height.



327

328 **Figure 8** The daily $PM_{2.5}$ distribution in the Yangtze River Delta (YRD, year 2015 day 25) and PRD (year 2015 day 26), with the
 329 TAP 1-km $PM_{2.5}$ predictions, the prediction with the historical land use predictors of year 2000, and the TAP 0.1-degree $PM_{2.5}$
 330 predictions. Gray shows the water body.

331 4 DISCUSSION

332 In this study, we fused the daily 10-km $PM_{2.5}$ predictions with satellite retrievals and land use data by a random forest model
 333 in the TAP framework and produced the open access daily average $PM_{2.5}$ distribution at a 1-km resolution with complete
 334 coverage from 2000 to the present. To improve the accuracy of the temporal variations in road distributions and other land use
 335 data, we processed the annual road map by fusing the OSM data with survey data and processed the annual population
 336 distribution by fusing the GPW and the WorldPop data. Our predictions showed an accuracy comparable with previous high-
 337 resolution $PM_{2.5}$ predictions, and our data were advantaged with complete coverage, time-varying land use predictors, and long
 338 temporal coverage. Compared to previous TAP products at approximately 10-km resolution, this 1-km resolution $PM_{2.5}$ data
 339 product revealed more local-scale spatial characteristics of the $PM_{2.5}$ distribution in China.

340 We conducted various evaluation analyses to optimize the model structure and select the appropriate predictors. When
 341 constructing the model structure and selecting the predictors, we not only considered the prediction accuracy but also
 342 considered the computation time, data updating frequency and long-term data availability. For example, we did not include
 343 the POI data as predictors since we have no access to historical POI data in China, and there is no appropriate model to correctly
 344 predict POI distributions in previous publications. Including more land use variables will certainly improve the model accuracy;
 345 however, since historical data are unavailable, doing so will increase the uncertainty in historical predictions. Similarly,

346 regarding the selected spatial predictors, we constructed temporally continuous data record with various geostatistical methods
347 and improve the data quality by fusing data from various sources to reveal the temporal changes in land use. Including
348 temporally mismatched predictors for PM_{2.5} prediction leads to misleading spatial patterns, especially in China with
349 considerable social economic development in the paste decades (Fig. 8). Additionally, we did not include any spatial and
350 temporal trends estimated from measurements in the hindcast model that could significantly improve the model performance
351 statistics in the evaluations. The measurement-based spatiotemporal trends did not necessarily reflect the pollution distribution
352 in regions without monitors (Bai et al., 2022a). Since the major aim of data fusing methods is to estimate the PM_{2.5} variations
353 in regions without monitors, including measurement-based smoothing trends in space and time will hinder the achievement of
354 this goal. Since the predictor processing and modelling of 1-km resolution data is computationally expensive, the predictor
355 selection and model structure selection not only improved model robustness, but also allowed us to run a more efficient model
356 daily and support near real-time data updating.

357 Our model still has several limitations. First, although we improved the model prediction accuracy during high pollution events,
358 we still noticed an underestimation of PM_{2.5} levels. Several reasons could explain this underestimation. The AOD retrievals
359 tended to misclassify high aerosol loading as cloud cover, leading to missing AOD during haze events. The CTM also hardly
360 predicts high pollution events. The missing satellite retrievals together with the underestimated CTM simulations resulted in
361 the underestimation of pollution levels. We noticed that all the predictors in the model are associated with some uncertainties
362 and these uncertainties together with the modelling error contributed to the uncertainties of the final PM_{2.5} predictions. Thus,
363 the quantification of the model uncertainties and their sources could be difficult. Here we conducted various model
364 performance evaluations to illustrate the prediction uncertainties from different angles. We suggested the usage of the out of
365 bag evaluation results as an approximate of the uncertainty of PM_{2.5} prediction after 2013 (Table 1), when the ground PM_{2.5}
366 monitoring is available; and the usage of the temporally cross-validation results as an approximate of the uncertainty before
367 2013 (Figure 4). Second, although we included some regional monitors to increase the density of monitors for model training,
368 the number of monitors in western China is still insufficient. Thus, the uncertainty of PM_{2.5} predictions in these regions lacking
369 data could be larger than in the regions with dense data. However, the distribution of monitors in China basically followed the
370 population distribution in which populous regions hold more monitors; thus, the key regions of air pollution control have more
371 training data and high-quality predictions, benefiting air quality management and environmental health studies in the future.

372 **5. Conclusions**

373 In this study, we constructed a high-resolution PM_{2.5} prediction model fused MAIAC satellite aerosol optical depth retrievals,
374 10-km TAP PM_{2.5} data, and land use variables including road map, population distribution, artificial impervious area, and
375 vegetation index. To describe the significant temporal variations in land use characteristics resulted from the economic
376 development in China, we constructed temporal continuous land use predictors through statistical and spatial modelling.
377 Optimization of model structure and predictors was conducted with various performance evaluation methods to balance the

378 model performance and computing cost. We revealed changes in local scale spatial pattern of PM_{2.5} associated with pollution
379 control measures. For example, pollution hotspots transferred from city centers to rural regions with limited air quality
380 monitoring. We showed that the land use data affected the predicted spatial distribution of PM_{2.5} and the usage of updated
381 spatial data is beneficial. The gridded 1-km resolution PM_{2.5} predictions can be openly accessed through the TAP website
382 (<http://tapdata.org.cn/>).

383 **Data availability**

384 The 1-km resolution PM_{2.5} predictions are available on the TAP website (<http://tapdata.org.cn/>).

385 **Author contribution**

386 QX and GG designed the study. SL ran the CMAQ simulations. LJ conducted the analysis on meteorological effects on the
387 haze events. XM collected and provided the road maps. QX trained the PM_{2.5} prediction models and conducted the
388 spatiotemporal analyses. QX prepared the manuscript with contributions from all co-authors.

389 **Competing interests**

390 The authors declare that they have no conflict of interest.

391 **Financial support**

392 This work was supported by the National Natural Science Foundation of China (grant no. 42007189, 42005135, and 41921005).

393 **Reference**

- 394 Bai, K., Li, K., Guo, J., and Chang, N.-B.: Multiscale and multisource data fusion for full-coverage PM_{2.5} concentration mapping: Can
395 spatial pattern recognition come with modeling accuracy?, *ISPRS Journal of Photogrammetry and Remote Sensing*, 184, 31-44,
396 <https://doi.org/10.1016/j.isprsjprs.2021.12.002>, 2022a.
- 397 Bai, K., Li, K., Ma, M., Li, K., Li, Z., Guo, J., Chang, N. B., Tan, Z., and Han, D.: LGHAP: the Long-term Gap-free High-resolution Air
398 Pollutant concentration dataset, derived via tensor-flow-based multimodal data fusion, *Earth Syst. Sci. Data*, 14, 907-927, 10.5194/essd-14-
399 907-2022, 2022b.
- 400 Bai, Z., Wang, J., Wang, M., Gao, M., and Sun, J.: Accuracy Assessment of Multi-Source Gridded Population Distribution Datasets in China,
401 *Sustainability*, 10, 1363, 10.3390/su10051363, 2018.
- 402 Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., and Worley, B. A.: LandScan: a global population database for estimating
403 populations at risk, *Photogrammetric engineering and remote sensing*, 66, 849-857, 2000.
- 404 Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., and Baptista, S. R.: Taking Advantage of the
405 Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4, *Papers in Applied Geography*, 1,
406 226-234, 10.1080/23754931.2015.1014272, 2015.

407 Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Xue, T., Tong, D., Zheng, B., Peng, Y., Huang, X., He, K., and Zhang, Q.:
408 Tracking Air Pollution in China: Near Real-Time PM_{2.5} Retrievals from Multisource Data Fusion, *Environmental Science & Technology*,
409 55, 12106-12115, 10.1021/acs.est.1c01863, 2021.

410 Goldberg, D. L., Gupta, P., Wang, K., Jena, C., Zhang, Y., Lu, Z., and Streets, D. G.: Using gap-filled MAIAC AOD and WRF-Chem to
411 estimate daily PM_{2.5} concentrations at 1 km resolution in the Eastern United States, *Atmospheric Environment*, 199, 443-452,
412 <https://doi.org/10.1016/j.atmosenv.2018.11.049>, 2019.

413 Haklay, M.: How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets,
414 *Environment and Planning B: Planning and Design*, 37, 682-703, 10.1068/b35097, 2010.

415 He, Q. Q., and Huang, B.: Satellite-based mapping of daily high-resolution ground PM_{2.5} in China via space-time regression modeling,
416 *REMOTE SENSING OF ENVIRONMENT*, 206, 72-83, 10.1016/j.rse.2017.12.018, 2018.

417 Hu, H. D., Hu, Z. Y., Zhong, K. W., Xu, J. H., Zhang, F. F., Zhao, Y., and Wu, P. H.: Satellite-based high-resolution mapping of ground-
418 level PM_{2.5} concentrations over East China using a spatiotemporal regression kriging model, *SCIENCE OF THE TOTAL ENVIRONMENT*,
419 672, 479-490, 10.1016/j.scitotenv.2019.03.480, 2019.

420 Huang, C., Hu, J., Xue, T., Xu, H., and Wang, M.: High-Resolution Spatiotemporal Modeling for Ambient PM_{2.5} Exposure Assessment in
421 China from 2013 to 2019, *Environmental Science & Technology*, 55, 2152-2162, 10.1021/acs.est.0c05815, 2021.

422 Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A. C., Schwartz, J., and Broday, D. M.: Estimating daily PM_{2.5} and
423 PM₁₀ across the complex geo-climate region of Israel using MAIAC satellite-based AOD data, *Atmospheric Environment*, 122, 409-416,
424 2015.

425 Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., and Korokin, S.: Multiangle implementation of atmospheric correction (MAIAC): 1.
426 Radiative transfer basis and look - up tables, *Journal of Geophysical Research: Atmospheres*, 116, 2011a.

427 Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korokin, S., Remer, L., Levy, R., and Reid, J.: Multiangle implementation of atmospheric
428 correction (MAIAC): 2. Aerosol algorithm, *Journal of Geophysical Research: Atmospheres*, 116, 2011b.

429 Lyapustin, A.: MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) Data User's Guide, 2018.

430 Ma, R. M., Ban, J., Wang, Q., Zhang, Y. Y., Yang, Y., Li, S. S., Shi, W. J., Zhou, Z., Zang, J. W., and Li, T. T.: Full-coverage 1 km daily
431 ambient PM_{2.5} and O₃ concentrations of China in 2005-2017 based on a multi-variable random forest model, *EARTH SYSTEM SCIENCE*
432 *DATA*, 14, 943-954, 10.5194/essd-14-943-2022, 2022.

433 Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., and Schipper, A. M.: Global patterns of current and future road infrastructure,
434 *Environmental Research Letters*, 13, 064006, 10.1088/1748-9326/aabd42, 2018.

435 Reed, F. J., Gaughan, A. E., Stevens, F. R., Yetman, G., Sorichetta, A., and Tatem, A. J.: Gridded Population Maps Informed by Different
436 Built Settlement Products, *Data*, 3, 10.3390/data3030033, 2018.

437 Wang, X. M., Xin, L. I., and Ming-Guo, M. A.: Advance and Case Analysis in Population Spatial Distribution Based on Remote Sensing
438 and GIS, *Remote Sensing Technology & Application*, 2004.

439 Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V., and Tatem, A.
440 J.: Spatially disaggregated population estimates in the absence of national population and housing census data, *Proceedings of the National*
441 *Academy of Sciences*, 115, 3529-3537, 10.1073/pnas.1715305115, 2018.

442 Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., and Cribb, M.: Estimating 1-km-resolution PM_{2.5} concentrations across China using
443 the space-time random forest approach, *Remote Sensing of Environment*, 231, 111221, <https://doi.org/10.1016/j.rse.2019.111221>, 2019.

444 Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., Lyapustin, A., Liu, L., Wu, H., and Song, Y.: Improved 1
445 km resolution PM_{2.5} estimates across China using enhanced space-time extremely randomized trees, *Atmos. Chem. Phys.*, 20, 3273-3289,
446 10.5194/acp-20-3273-2020, 2020.

447 Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T., and Cribb, M.: Reconstructing 1-km-resolution high-quality PM_{2.5} data
448 records from 2000 to 2018 in China: spatiotemporal variations and policy implications, *Remote Sensing of Environment*, 252, 112136,
449 <https://doi.org/10.1016/j.rse.2020.112136>, 2021.

450 Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A., and Liu, Y.: Full-coverage high-resolution daily PM_{2.5} estimation
451 using MAIAC AOD in the Yangtze River Delta of China, *Remote Sensing of Environment*, 199, 437-446, 2017.

452 Xiao, Q., Geng, G., Liang, F., Wang, X., Lv, Z., Lei, Y., Huang, X., Zhang, Q., Liu, Y., and He, K.: Changes in spatial patterns of PM_{2.5}
453 pollution in China 2000-2018: Impact of clean air policies, *Environment International*, 141, 105776,
454 <https://doi.org/10.1016/j.envint.2020.105776>, 2020.

455 Xiao, Q., Geng, G., Cheng, J., Liang, F., Li, R., Meng, X., Xue, T., Huang, X., Kan, H., Zhang, Q., and He, K.: Evaluation of gap-filling
456 approaches in satellite-based daily PM_{2.5} prediction models, *Atmospheric Environment*, 244, 117921,
457 <https://doi.org/10.1016/j.atmosenv.2020.117921>, 2021a.

458 Xiao, Q., Geng, G., Xue, T., Liu, S., Cai, C., He, K., and Zhang, Q.: Tracking PM_{2.5} and O₃ Pollution and the Related Health Burden in
459 China 2013-2020, *Environmental Science & Technology*, 10.1021/acs.est.1c04548, 2021b.

460 Xiao, Q., Zheng, Y., Geng, G., Chen, C., Huang, X., Che, H., Zhang, X., He, K., and Zhang, Q.: Separating emission and meteorological
461 contributions to long-term PM_{2.5} trends over eastern China during 2000-2018, *Atmos. Chem. Phys.*, 21, 9475-9496, 10.5194/acp-21-9475-
462 2021, 2021c.

463 Xie, Y. Y., Wang, Y. X., Zhang, K., Dong, W. H., Lv, B. L., and Bai, Y. Q.: Daily Estimation of Ground-Level PM2.5 Concentrations over
464 Beijing Using 3 km Resolution MODIS AOD, ENVIRONMENTAL SCIENCE & TECHNOLOGY, 49, 12280-12288,
465 10.1021/acs.est.5b01413, 2015.
466 Zhang, T., Zhu, Z., Gong, W., Zhu, Z., Sun, K., Wang, L., Huang, Y., Mao, F., Shen, H., Li, Z., and Xu, K.: Estimation of ultrahigh resolution
467 PM2.5 concentrations in urban areas using 160 m Gaofen-1 AOD retrievals, Remote Sensing of Environment, 216, 91-104,
468 <https://doi.org/10.1016/j.rse.2018.06.030>, 2018.
469 Zhang, Y., Li, X., Wang, A., Bao, T., and Tian, S.: Density and diversity of OpenStreetMap road networks in China, Journal of Urban
470 Management, 4, 135-146, <https://doi.org/10.1016/j.jum.2015.10.001>, 2015.

471