Sequential spatiotemporal distribution of $PM_{2.5}$, SO_2 and Ozone in China from 2015 to 2020

Yufeng Chi [a], Yu Zhan [b], Kai Wang [c], Hong Ye [d, e, f, g, *]

a. School of Information Engineering, Sanming University, Sanming 365004, China
b. Department of Environmental Science and Engineering, Sichuan University, Chengdu 610065, China
c. China-UK Low Carbon College, Shanghai Jiaotong University, Shanghai 200000, China
d. Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
e. Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
f. CAS Haixi Industrial Technology Innovation Center in Beilun, Ningbo 315800, China
g. Xiamen Key Laboratory of smart management of urban environment Xiamen 361021, China
Correspondence to: Hong Ye (hye@iue.ac.cn)

To supplement the content of the main article, the Supplementary Section includes Data S1, Data S2 and Figures S3-S6.

Data S1

Data preprocessing includes remote sensing data gap filling and remote sensing data grid remaking. Missing filling of remote sensing data adopts the TWS method developed by Chi et al. Remote sensing data regrid includes spatial regrid and time regrid filling.

1. TWS method

We use the previously developed TWS method to fill the remote sensing data gaps in this study (Chi et al., 2020). In addition, we also successfully assisted in the development of the RF-RID model using TWS in chi et al. (Chi and Zhan, 2022), and applied it to regional NO_2 spatiotemporal simulations. The specific steps and equations of TWS can refer to the above two manuscripts. MAIAC AOD uses the first step and iterative second step of TWS, the OMI SO and OMI O_3 column is recovered through the iterative second step of TWS.

2. Spatiotemporal Data Grid Rework Overview

Transform the multi-source data to form a consistent temporal and spatial resolution dataset with a temporal resolution of 1 day and a spatial resolution with the WGS grid coordinate system and resolution in the auxiliary data. Elevation (Ele) and slope (SLOP) are calculated by SRTM. Ele, SLOP, Land Use (LUCC), Road Data (RL), MAIAC AOD, NDVI, Population (POP), AHI AOD, OMI SO and OMI O_3 overlaid on WGS grid, based on target data and Coverage area ratio of WGS grid pixels, calculated using weighted average of reproduced target data. Next, populate the data (population data, land use, NDVI, etc.) at a time resolution greater than one day. For example, NDVI fills gaps in the time range with NDVI values based on a time resolution of 16 days. Finally, the meteorological parameters (temperature (Tem), pressure (Pr), humidity (Hum) and wind speed (Ws)) were interpolated using the spline method.

Data S2

We collected the air pollutant monitoring data of $PM_{2.5}$, SO_2, and O_3 from 2015 to 2020 and the remote sensing data corresponding to the spatial location of air pollutant stations. There are a total of 4,405,312 valid data, and the histogram distribution of the data was analyzed (Figure S1). Monthly average of ground monitoring of air pollution in China (Figure S2). The annual average of air pollution ground monitoring in China is located in Table S1.

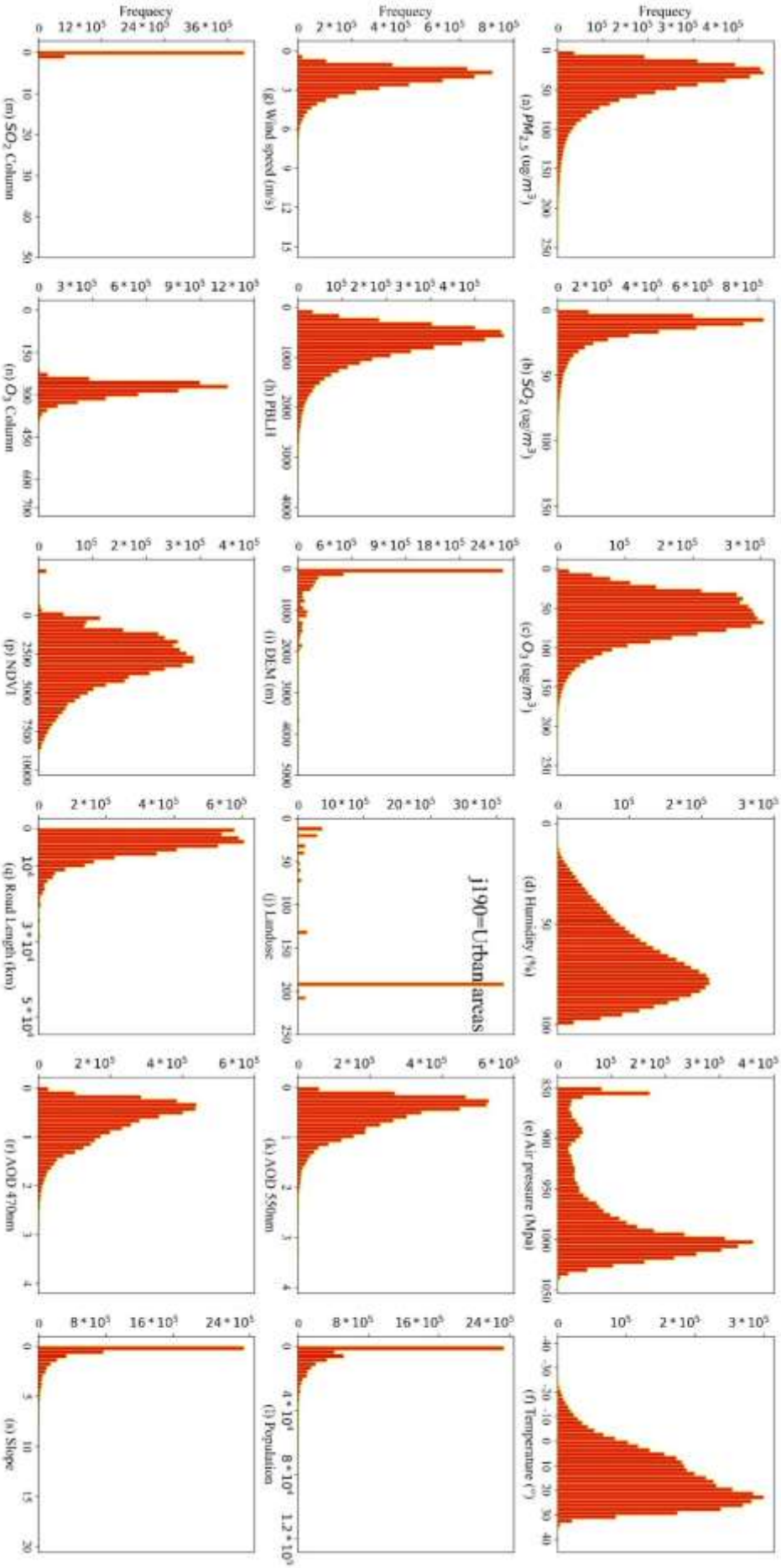
Frequecy
(a) $PM_{2.5}$ (ug/m^3)
(g) Wind speed (m/s)
(m) SO_2 Column
(b) SO_2 (ug/m^3)
(h) PBLH
(n) O_3 Column
(c) O_3 (ug/m^3)
(i) DEM (m)
(p) NDVI
(d) Humidity (%)
(j) Landuse
j190=Urban areas
(q) Road Length (km)
(e) Air pressure (Mpa)
(k) AOD 550nm
(r) AOD 470nm
(f) Temperature (°)
(l) Population
(s) Slope

Figure S1 Factor histogram. Where the one with the highest LandUse represents the urban area. It shows that most air pollutant monitoring stations are mainly located in urban areas.

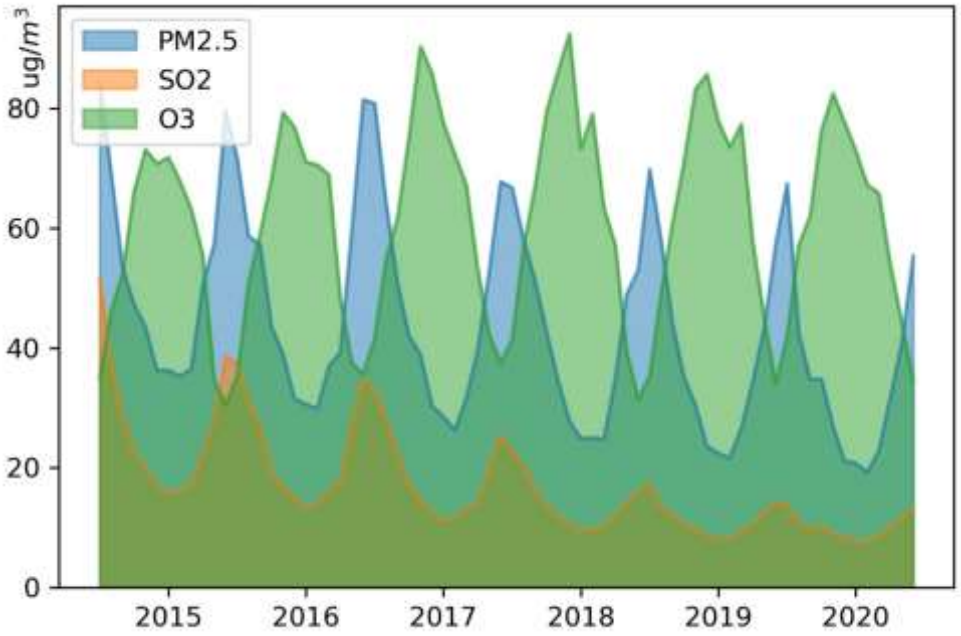

Figure S2. Monthly average of ground monitoring of air pollution in China.

Table S1 The annual average of air pollution ground monitoring in China

Year	$PM_{2.5}$(ug/m^3)	SO_2(ug/m^3)	O_3(ug/m^3)
2015	52.46	26.15	55.57
2016	48.26	22.45	58.45
2017	46.14	18.49	63.28
2018	41.04	13.89	63.92
2019	38.88	11.15	62.55
2020	34.84	10.02	61.38

Figure S3 Correlation coefficients of $PM_{2.5}$, SO_2, and O_3 and other elements at the level of ground observation sites.

From Figure S1, the distributions of different air pollutant concentrations are relatively close, with all unimodal skewed distributions, and most of the remaining parameters are also skewed. Combined with the correlation heat map of Figure S3, it can be found that $PM_{2.5}$ is positively correlated with SO_2 (r=0.44), and O_3 is negatively correlated with $PM_{2.5}$ and SO_2, respectively (R is -0.22 and -0.18, respectively). There is also a certain degree of correlation between the object and other selected parameters.

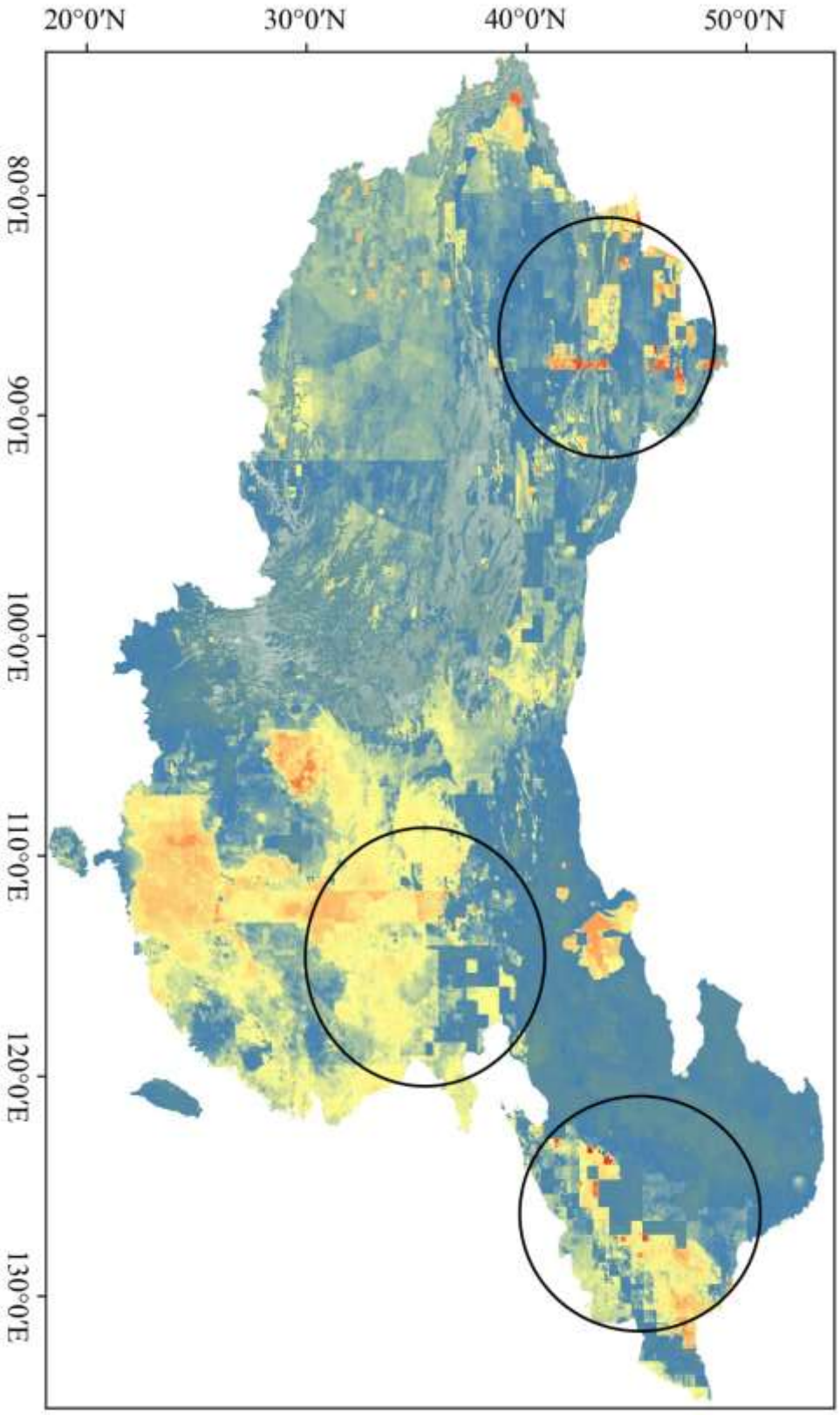

Figure S4 Schematic diagram of the “bands” or “patches” phenomenon simulated in the spatial distribution of air pollutants. The black marks clearly show this phenomenons.

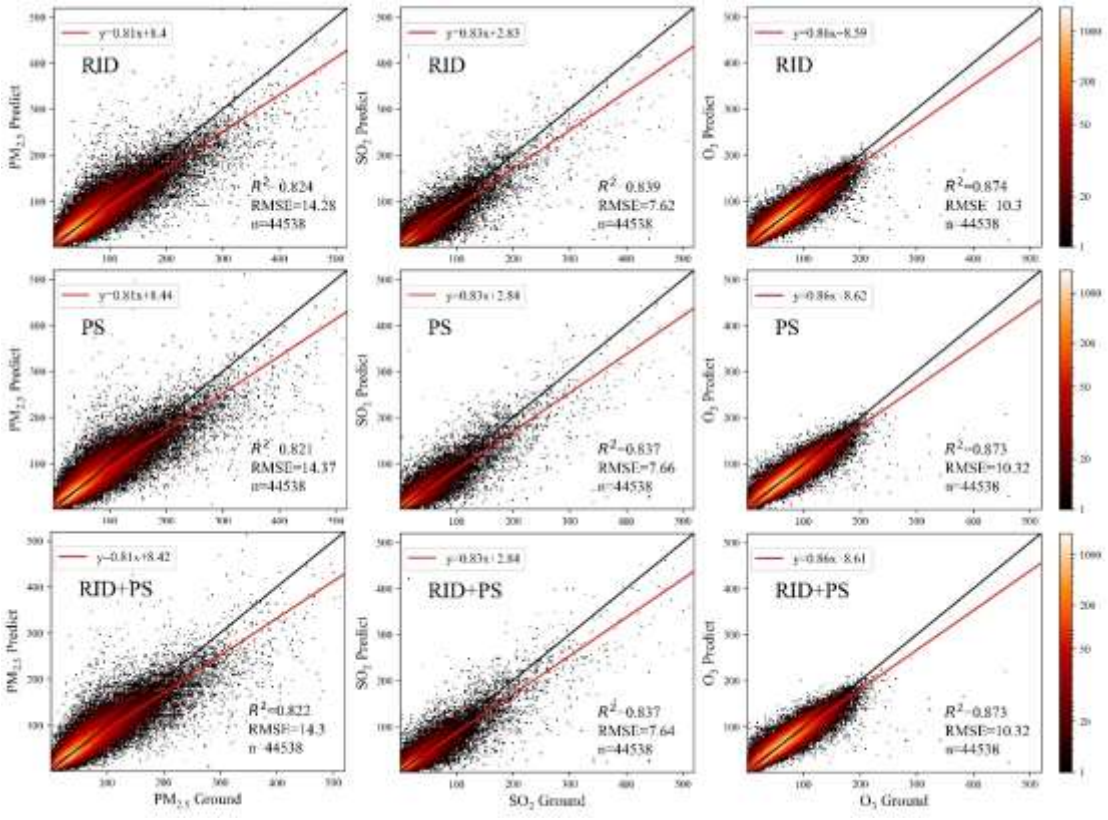

Figure S5 LightGBM added RID, Ps and RID+Ps parameters to the CV evaluation difference of $PM_{2.5}$, SO_2, and O_3. The R2 of RID is the highest and the RMSE is the lowest. Among them, the unit of RMSE is $\mu g/m^3$.

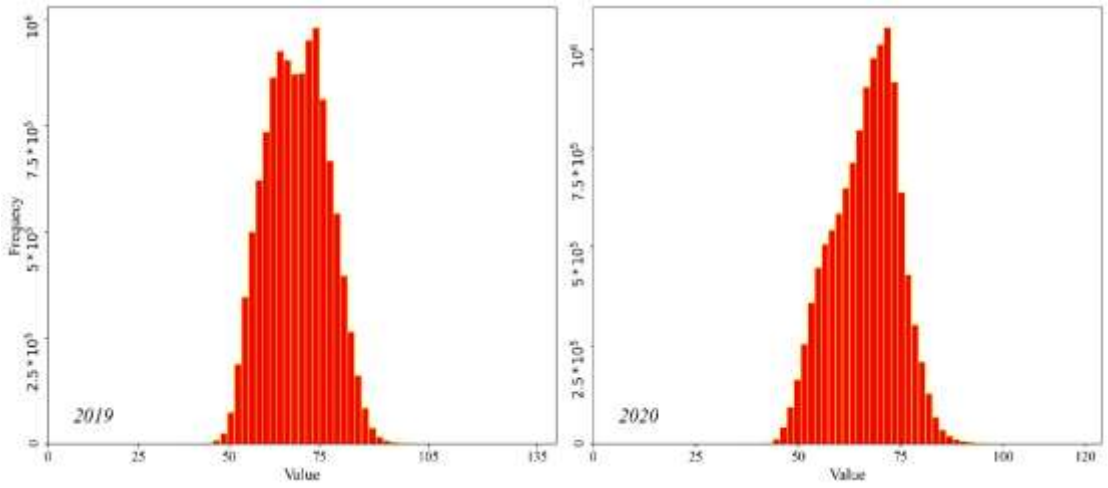

Figure S6 Histogram of API statistics for 2019 and 2020. The 2020 API results are more concentrated than 2019, while having smaller values.

Reference

Chi, Y. and Zhan, Y.: A Simple and Effective Random Forest Refit to Map the Spatial Distribution of NO2 Concentrations, 10.3390/atmos13111832, 2022.

Chi, Y., Wu, Z., Liao, K., and Ren, Y.: Handling Missing Data in Large-Scale MODIS AOD Products Using a Two-Step Model, Remote Sens., 12, 3786, 10.3390/rs12223786, 2020.