

# CHAP

ChinaHighAirPollutants

中国高分辨率高质量近地表空气污染物数据集

# CHAP



Jing Wei

Chinese

HOME

PUBLICATION

PRESENTATION

RESEARCH

PRODUCT

NEWS

## ChinaHighAirPollutants (CHAP)

**New update:** [Daily seamless 1 km PM<sub>x</sub> and composition data released!](#)

- **Brief Introduction**

The ChinaHighAirPollutants (CHAP) dataset refers to the **long-term, full-coverage, high-resolution**, and **high-quality** datasets of ground-level air pollutants for China. It is generated from the big data (e.g., ground-based measurements, satellite remote sensing products, atmospheric reanalysis, and model simulations) using artificial intelligence by considering the spatiotemporal heterogeneity of air pollution. The CHAP dataset contains **7** major air pollutants (i.e., **PM<sub>1</sub>**, **PM<sub>2.5</sub>**, **PM<sub>10</sub>**, **O<sub>3</sub>**, **NO<sub>2</sub>**, **SO<sub>2</sub>**, and **CO**), and **PM<sub>2.5</sub> compositions** (e.g., **SO<sub>4</sub><sup>2-</sup>**, **NO<sub>3</sub><sup>-</sup>**, **NH<sub>4</sub><sup>+</sup>**, **Cl<sup>-</sup>**, and **BC**, et al). This CHAP dataset is **public** and **freely** open to all users!

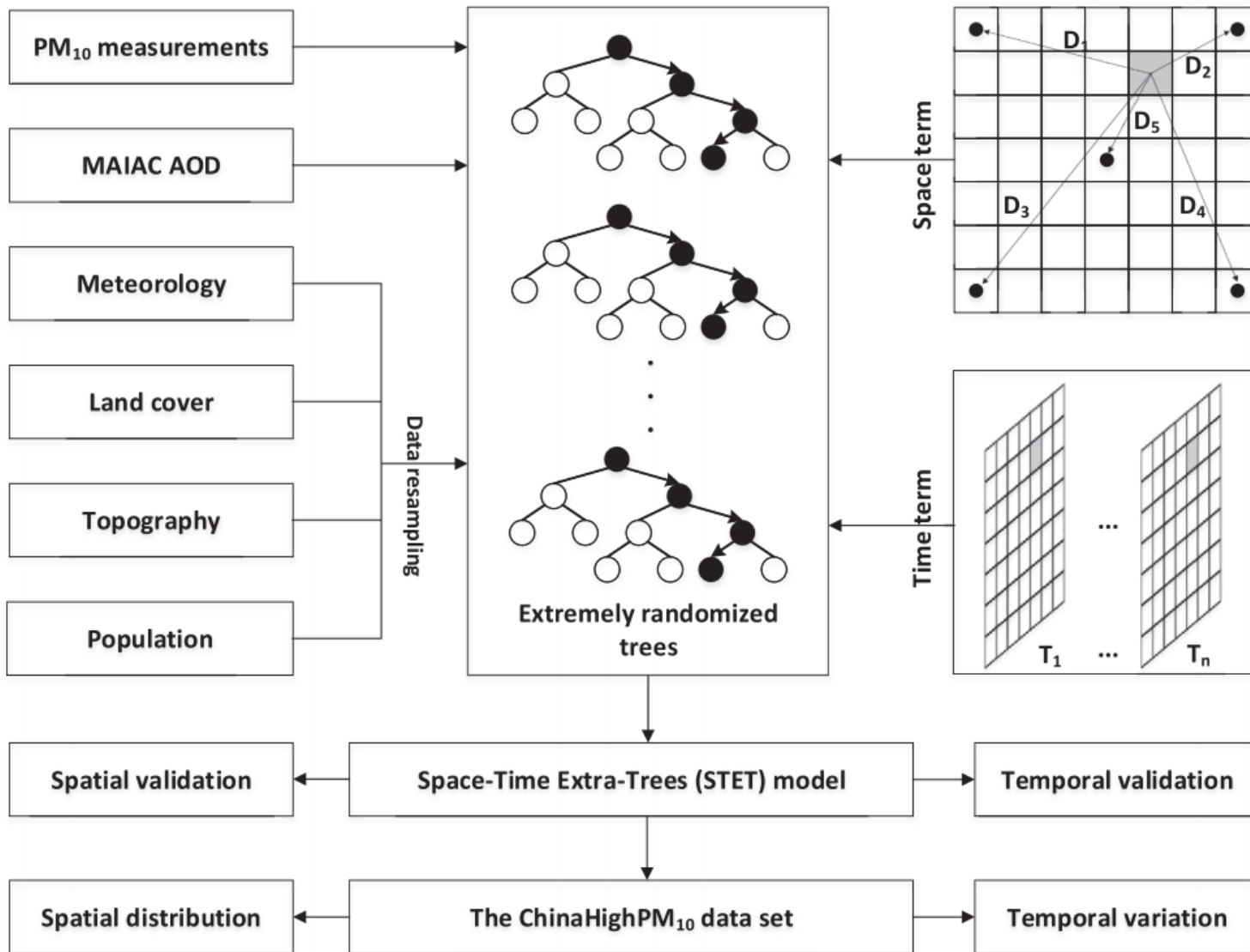
- **Dataset summary**

ChinaHighAirPollutants (CHAP)								
Air Pollutant	Main predictor	Spatial resolution	Missing values	Temporal resolution				Available period (yyyy/mm)
				Hourly	Daily	Monthly	Yearly	
PM <sub>1</sub>	Big data	1 km	No		√	√	√	2000/01 – 2021/12
PM <sub>2.5</sub>	Big data	1 km	No		√	√	√	2000/01 – 2021/12
	Himawari-8	5 km	Yes	√	√	√	√	2018/01 – 2018/12
PM <sub>10</sub>	Big data	1 km	No		√	√	√	2000/01 – 2021/12

Copyright © 2020 Jing Wei

Website: <https://weijing-rs.github.io/product.html>

# ChinaHighPM<sub>10</sub> (Method & Updates)



## Method

Space-Time Extra-Trees (STET) model

## Main updates (Version 4)

1) AOD gap filling:

Fill the AOD gaps from big data using machine learning

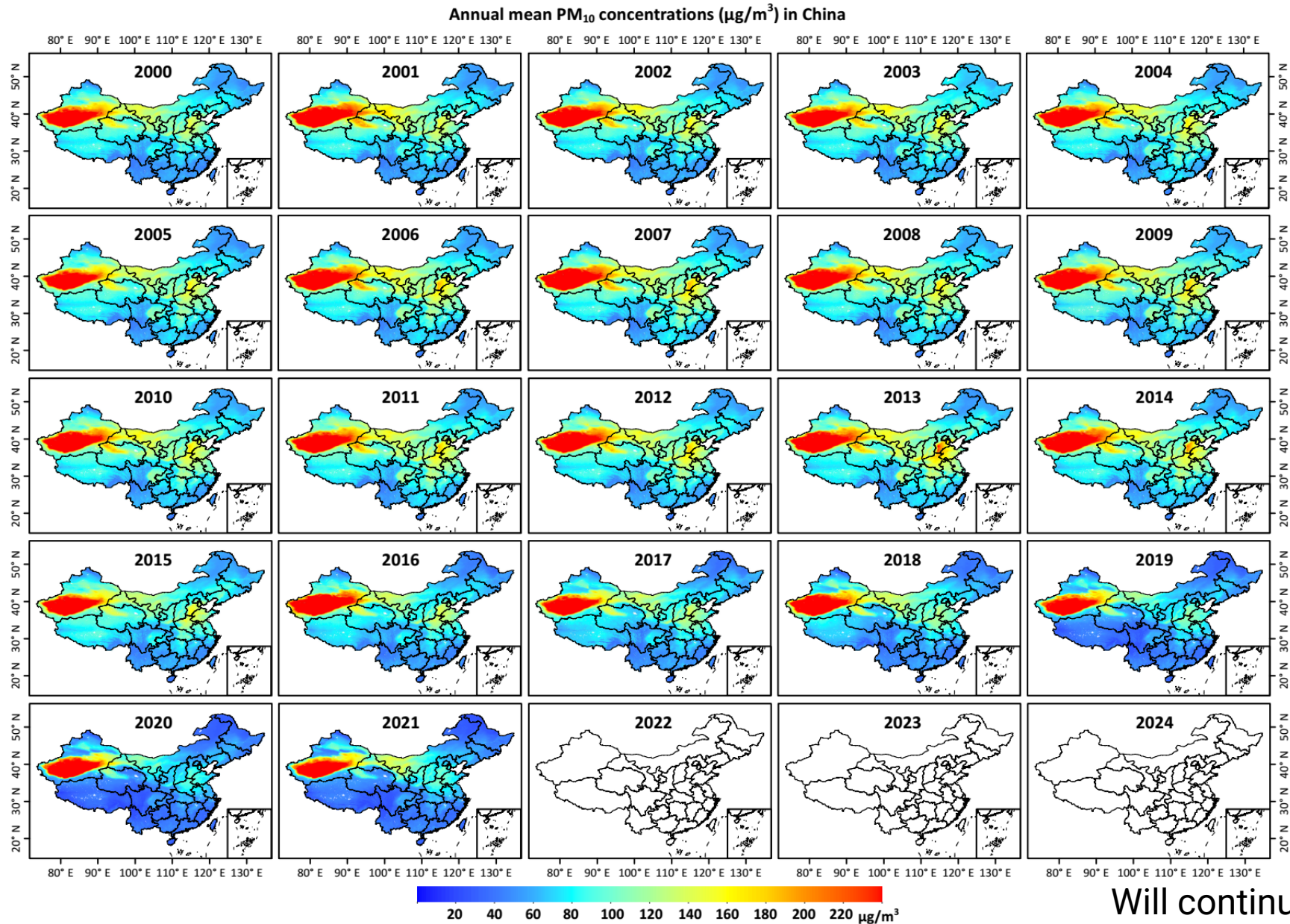
2) Update data sources

e.g., MERRA2 PM compositions

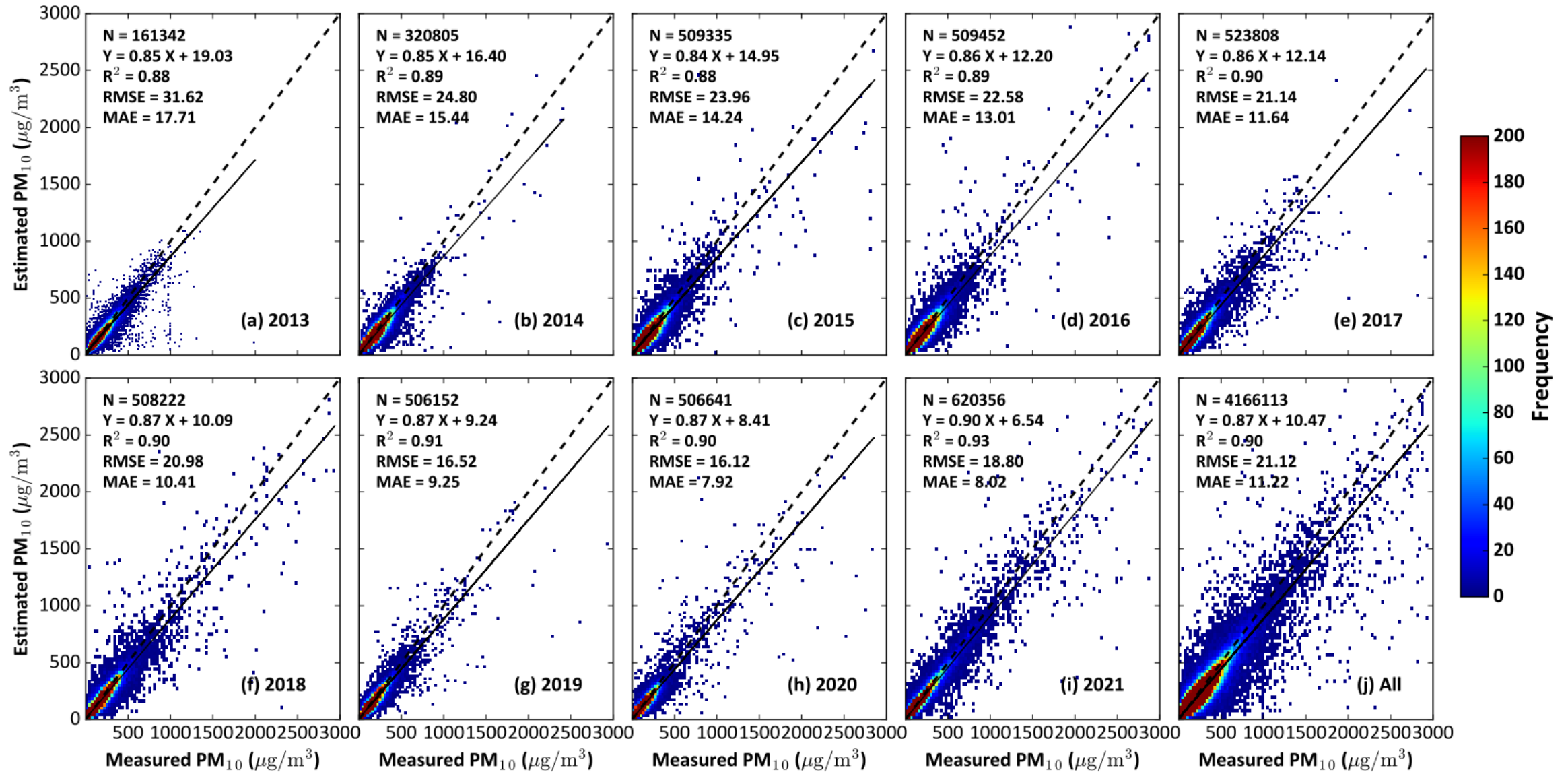
CAMS emission inventory

Flowchart of STET model (Wei et al., EI, 2021)

# ChinaHighPM<sub>10</sub> (1 km, 2000-2021, Version 4)

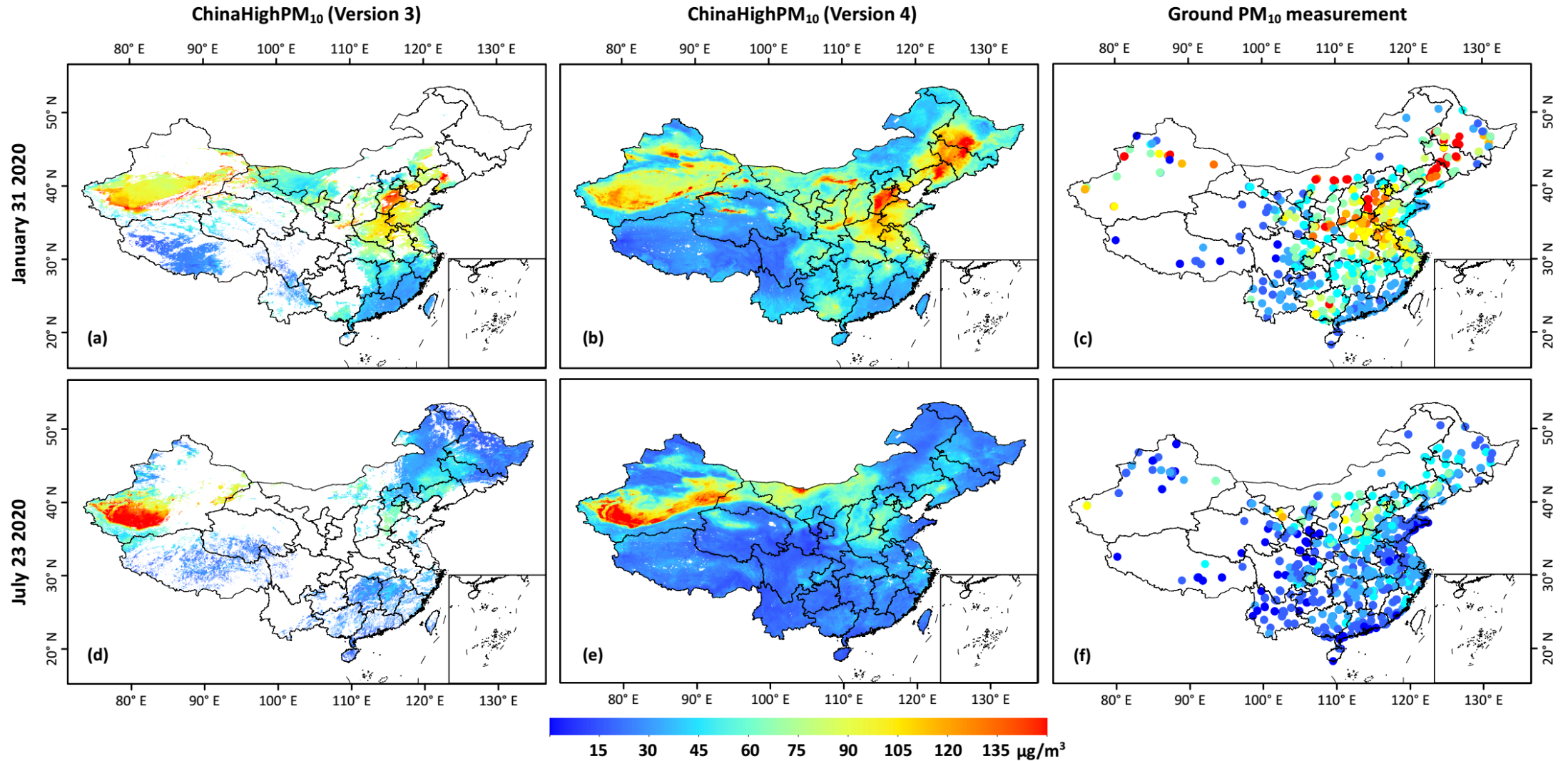


# 10-fold Cross Validation



Overall accuracy: CV- $R^2 = 0.90$ , RMSE =  $21.12 \mu g/m^3$

# Version Comparison (4 & 3)



The V4 data filled the missing values of satellite AOD products and provided **seamless** daily PM<sub>10</sub> concentrations across China, significantly improving the data availability by **60%** compared to the V3 dataset.

# ChinaHighPM<sub>10</sub> product (format)

Panoply: Panoply — Sources

File Edit View History Bookmarks Plot Window Help

Create Plot Combine Plot Open Dataset

Remove Remove All Hide Info

Name	Long Name	Type
CHAP_PM10_D1K_20200101_V4.nc	CHAP_PM10_D1K_20200101_V4.nc	Local File
lat	lat	1D
lon	lon	1D
PM10	PM10	Geo2D

PM10 in CHAP\_PM10\_D1K\_20200101\_V4

File Edit View History Bookmarks Plot Window Help

Plot Array 1

PM10

PM10 (µg/m3)

6.2 106.7 207.3 307.8 408.4 508.9

Data Min = 6.2, Max = 508.9, Mean = 68.0

Array(s) Scale Map Overlays Shading Contours Vectors Labels

Plot Map of Array 1 Only  Interpolate

Array 1: PM10

No additional dimensions

```
Variable "PM10"
In file "CHAP_PM10_D1K_20200101_V4.nc"
ushort PM10(lat=3571, lon=6148):
:units = "µg/m3";
:scale_factor = 0.1f; // float
:add_offset = 0.0f; // float
:_FillValue = 65535US; // ushort
:_ChunkSizes = 36U, 62U; // uint
```

Format: CHAP\_PM10\_D1K\_20200101\_V4.nc

**PM10:** ground-level PM<sub>10</sub>

**D1K:** Daily 1 km (**M1K:** Monthly 1 km; **Y1K:** Yearly 1 km)

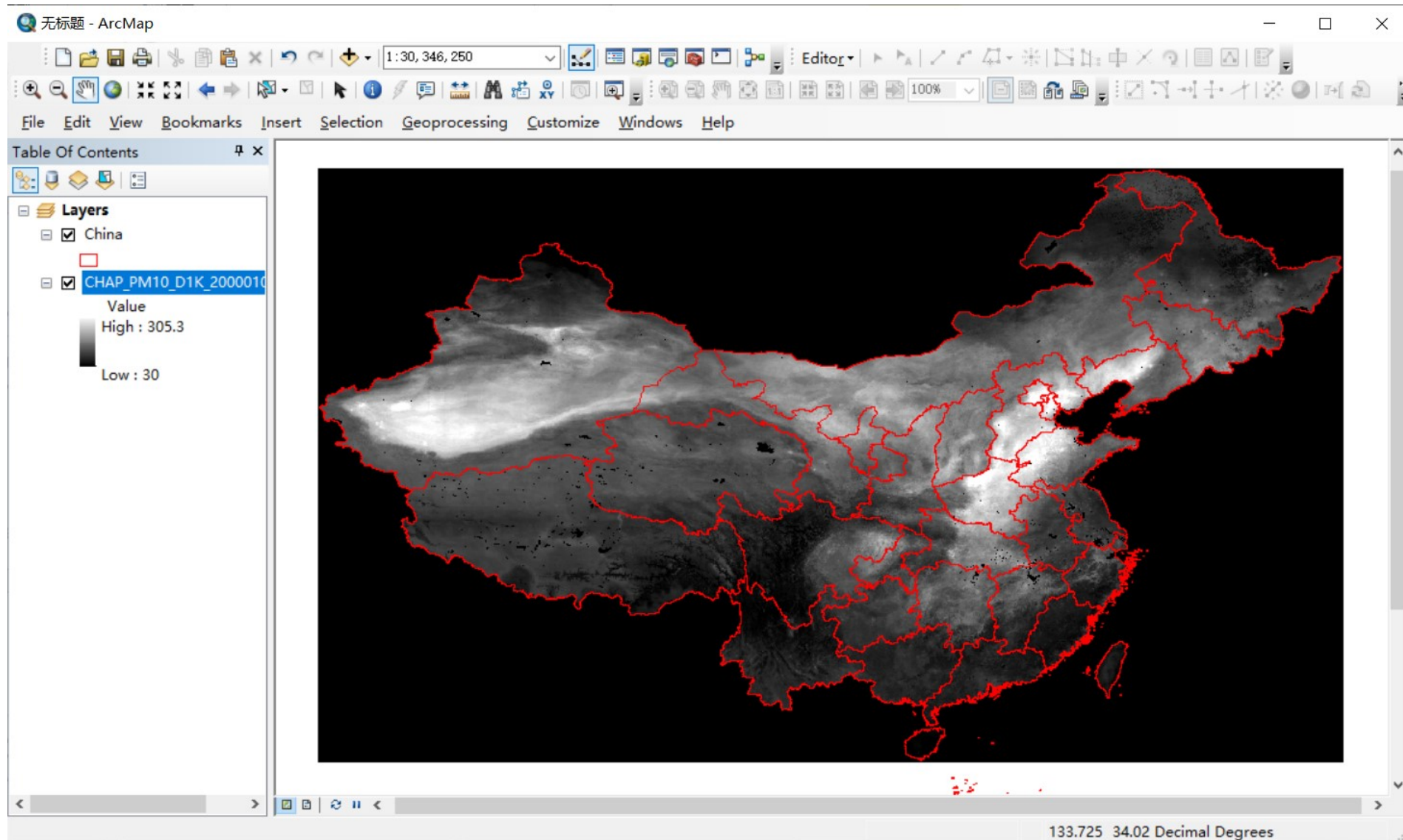
**20200101:** Date [Year, Month, Day]

**V4:** Version 4

Check the .nc file using **Panoply** or **HDF Explorer**

# How to read?

Use Python, Matlab, and IDL codes (nc2geotiff codes.rar) to batch convert NetCDF (.nc) to GeoTIFF (.tif).





# Reference

---

[1] Wei, J., Li, Z., Xue, W., Sun, L., Fan, T., Liu, L., Su, T., and Cribb, M. [The ChinaHighPM<sub>10</sub> dataset: generation, validation, and spatiotemporal variations from 2013 to 2019 across China](https://doi.org/10.1016/j.envint.2020.106290). *Environment International*, 2021, 146, 106290. <https://doi.org/10.1016/j.envint.2020.106290>

# Contact

---

If you use the ChinaHighPM<sub>10</sub> dataset for related study, please cite the corresponding reference (Wei et al., EI, 2021). **Note that this dataset is continuously updated, and if you need more data or have any questions, please contact me ([weijing\\_rs@163.com](mailto:weijing_rs@163.com); [weijing.rs@gmail.com](mailto:weijing.rs@gmail.com)).**