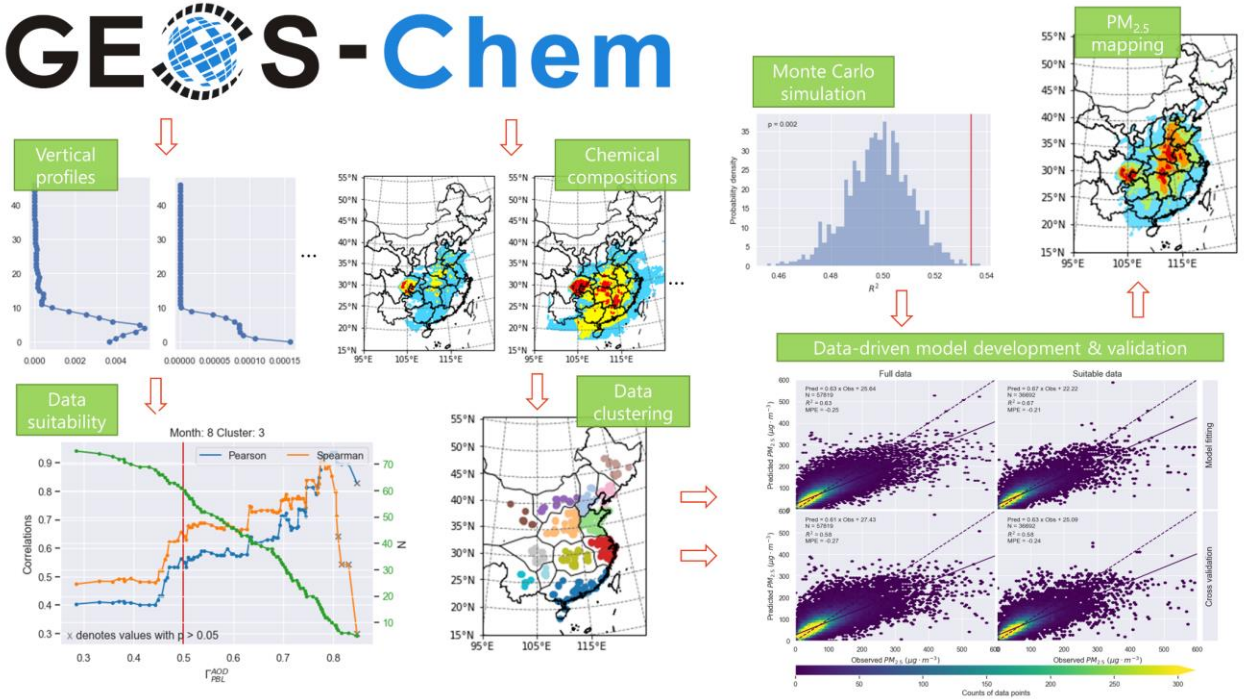
A model framework to reduce bias in ground-level $PM_{2.5}$ concentrations inferred from satellite-retrieved AOD

Fei Yao[1] (fei.yao@ed.ac.uk), Paul I. Palmer[1,2]

[1]School of GeoSciences, University of Edinburgh, UK;
[2]NCEO, University of Edinburgh, UK

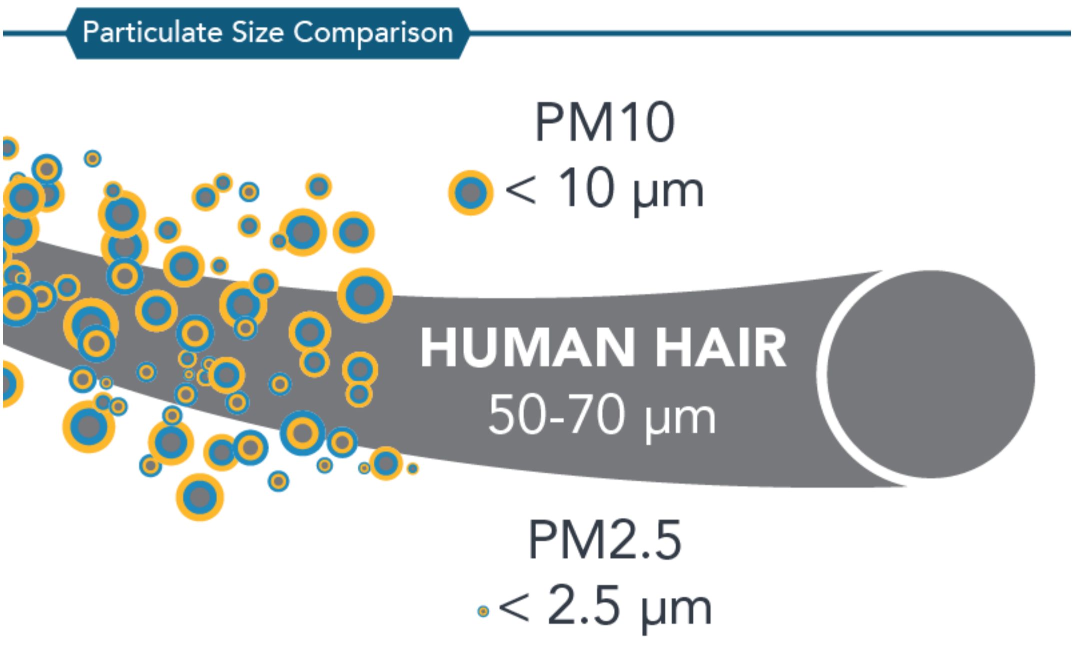
THE UNIVERSITY of EDINBURGH
School of GeoSciences



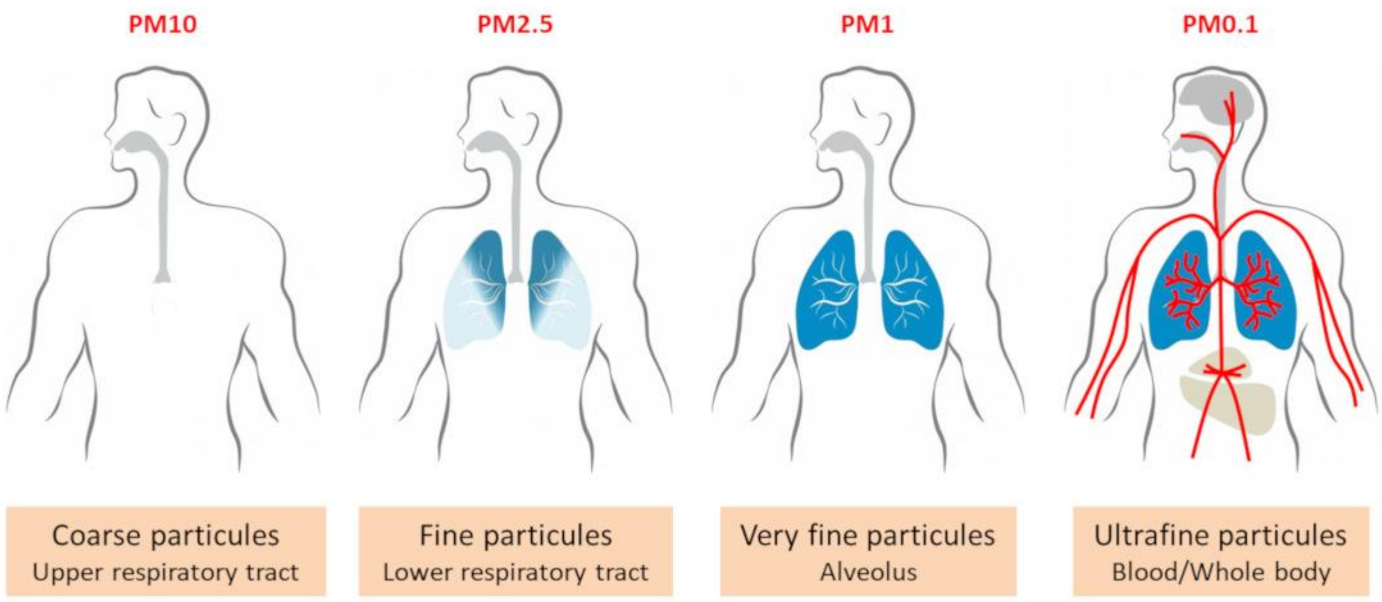
@北京大学蔚蓝空间大气科学研究生论坛第24期
2021年5月

What is $PM_{2.5}$?

- Particles with an aerodynamic dimeter less than 2.5 µm.
- Well-documented deleterious impacts on human health.



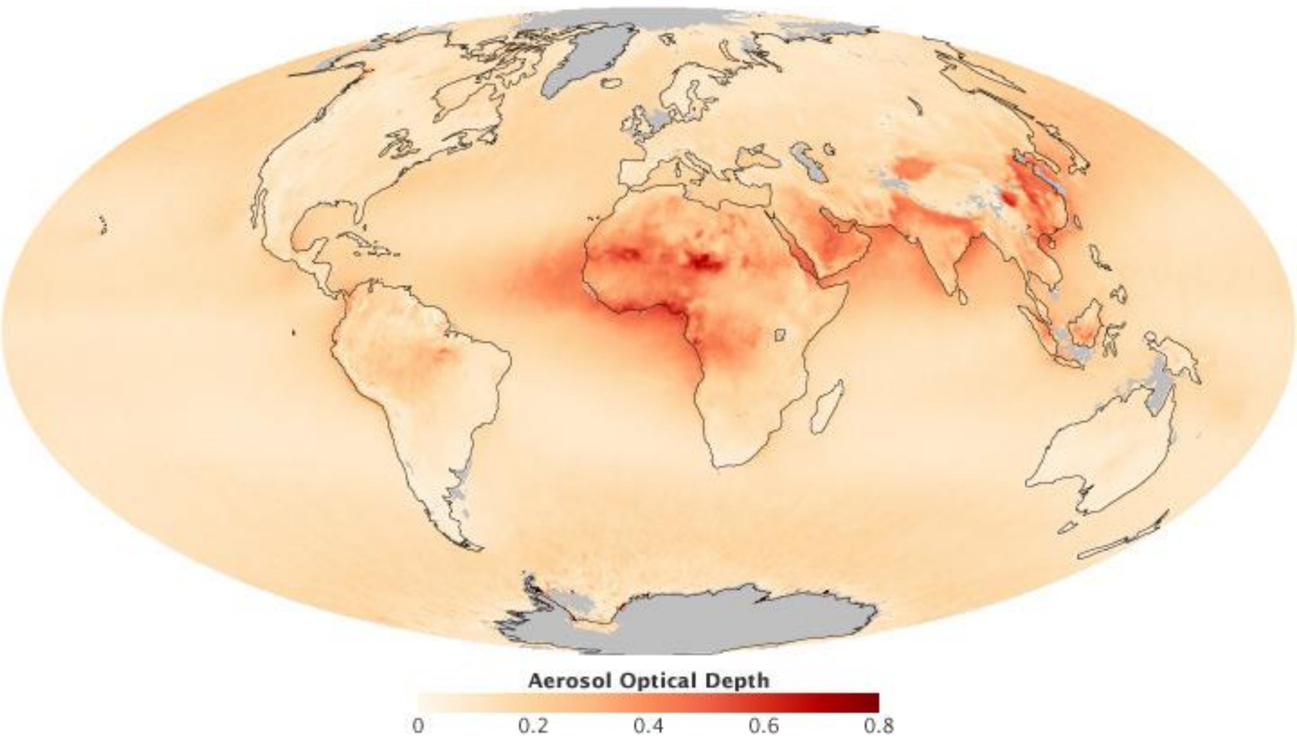
CARB



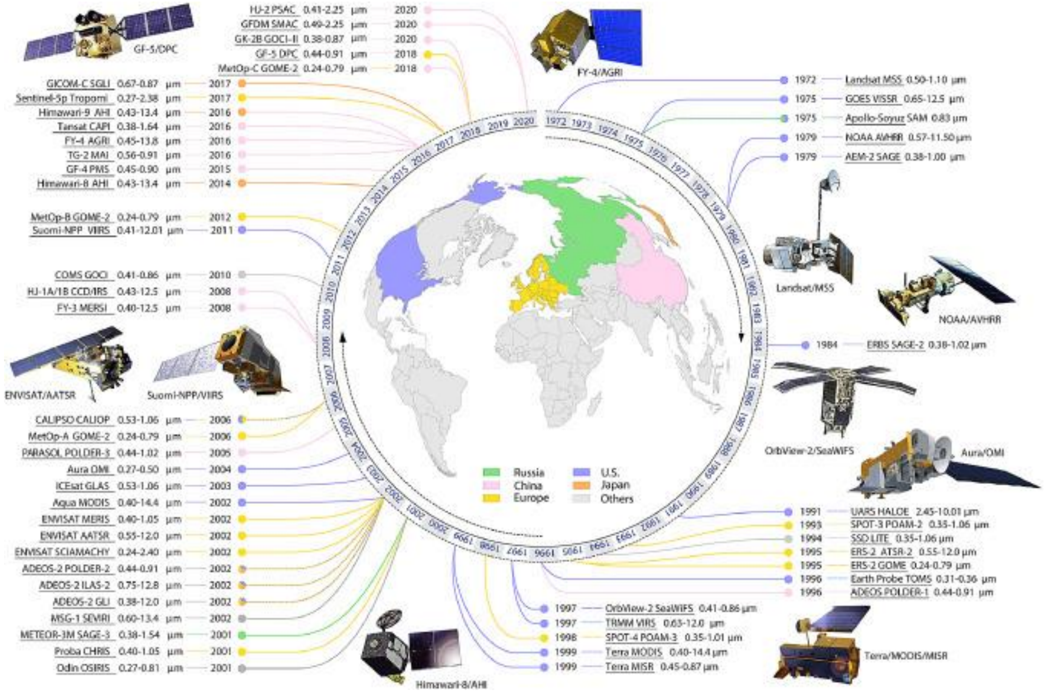
DANEL, 2019

What is AOD?

- Aerosol optical depth (AOD) is a measure of the amount of light that aerosols scatter and absorb in the atmosphere.
- Globally retrievable from satellite remote sensing observations.



NASA



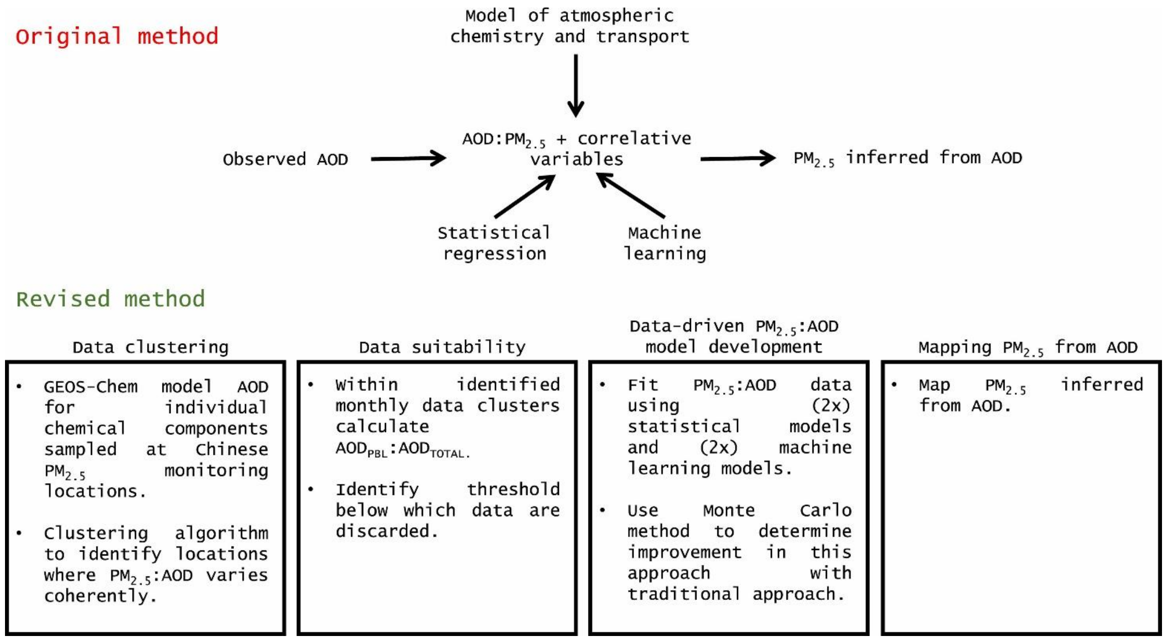
Zhang et al., 2021

Correlate $PM_{2.5}$ and AOD?

- While both indices describe the aerosols in the atmosphere, converting the latter to the former is non-trivial.

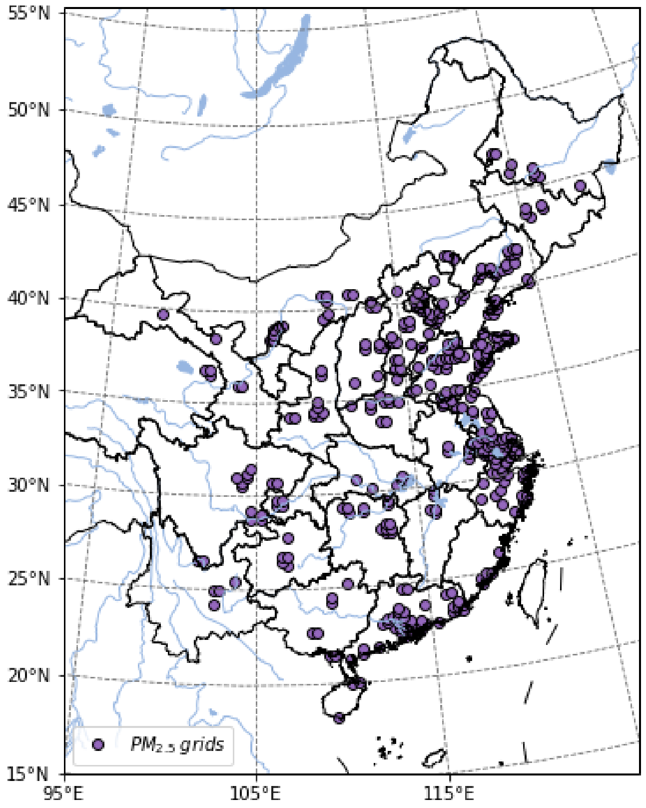
- Ground-level $PM_{2.5}$ values are mass concentrations of fine particles measured under controlled relative humidity conditions.

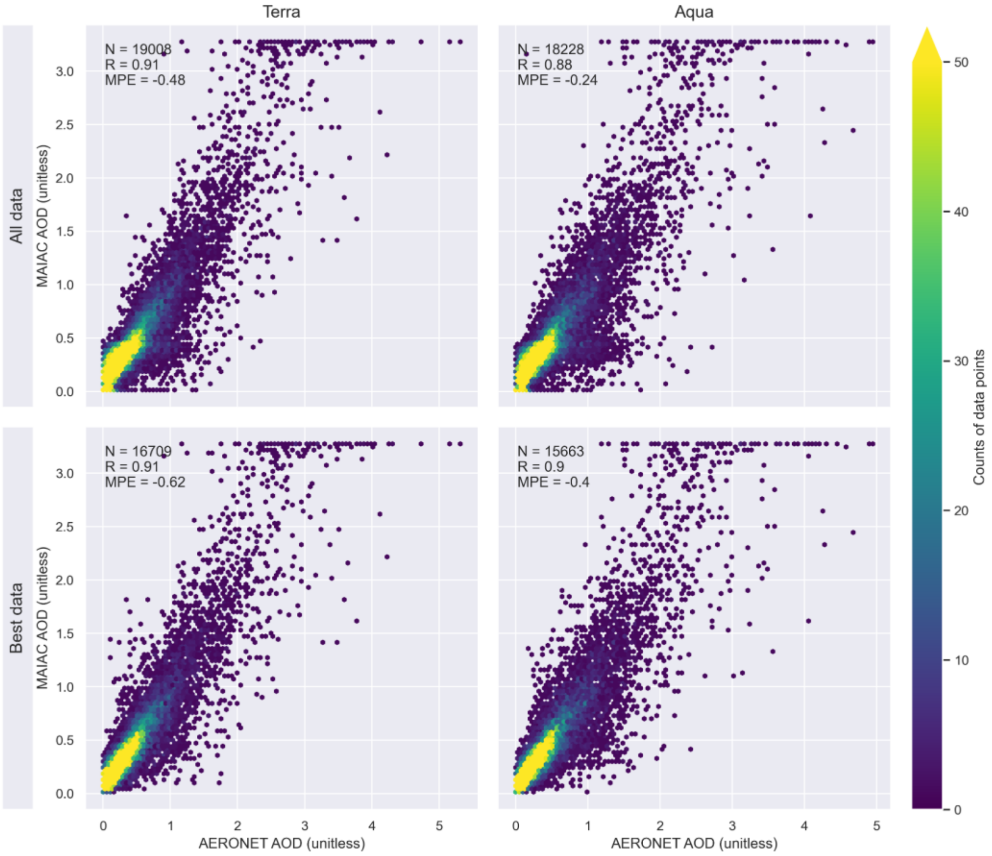
- AOD reflects a vertical integration of aerosol extinction coefficients at a specific wavelength.



$PM_{2.5}$ and AOD observations

- Study area limited to eastern China (95-140° E, 15-55° N).
- Likely erroneous values of $PM_{2.5}$ are removed following Jiang et al., 2020.
- All MODIS MAIAC AOD.





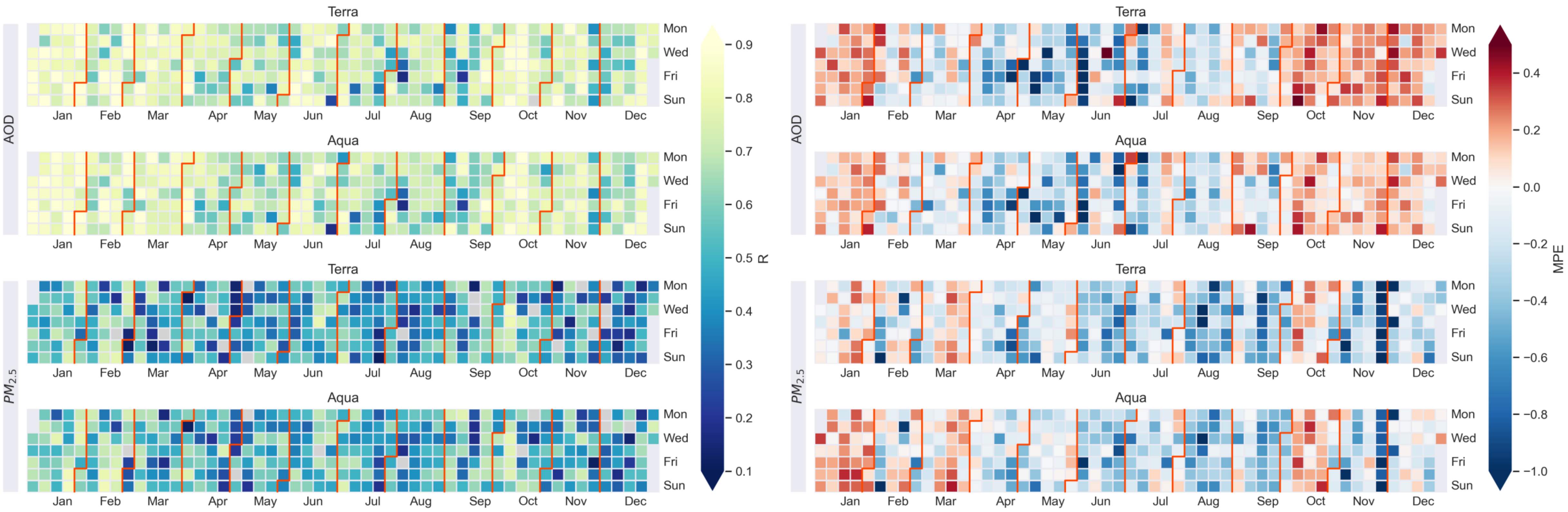
GEOS-Chem model configurations

GEOS-Chem v12.5.0

- 0.25x0.3125 simulations during 201310-201412 driven by GEOS-FP meteorology using full chemistry in the troposphere coupled with complex SOA and semi-volatile POA with the first three months as spin-up. Boundary conditions are taken from a self-consistent global version of model run.
- Emission inventories: Anthropogenic from MEIC (Zheng et al., 2018), Biogenic from MEGAN (Guenther et al., 2012), Pyrogenic from GFED (van der Warf et al., 2017), etc.
- Model outputs: 3-D fields of aerosol species including sulfate, nitrate, ammonium, POA, SOA, black carbon, dust, and sea salt. A linear sum with varying weights of the mass concentrations of these species leads to $PM_{2.5}$, while further combination with aerosol mass extinction coefficients gives AOD.

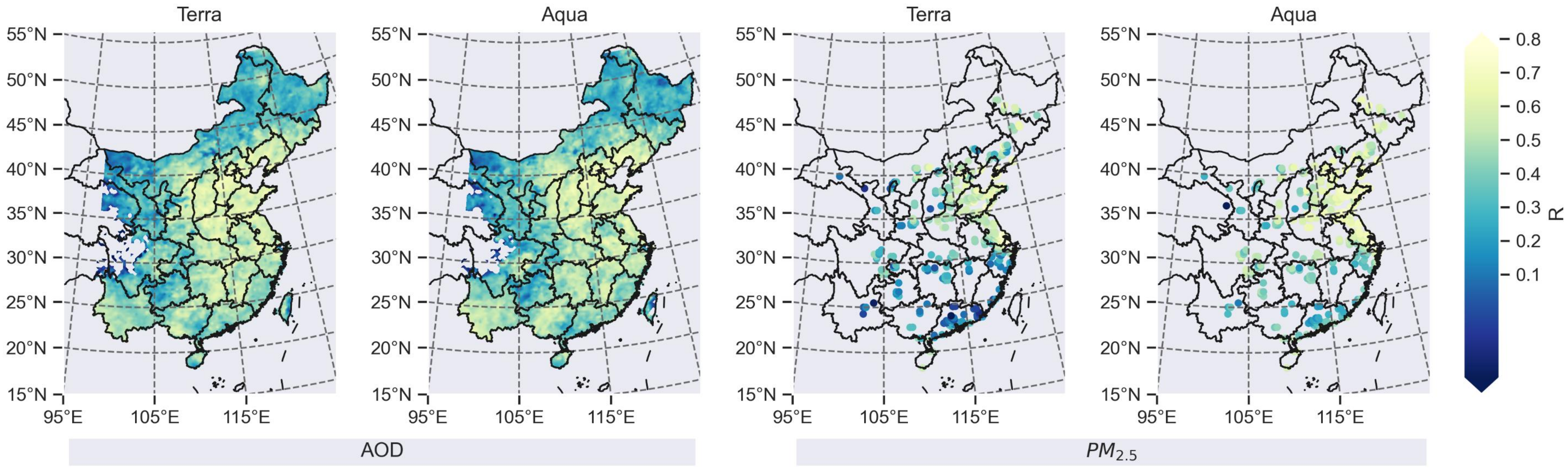
GEOS-Chem model evaluations

Temporal perspective against $PM_{2.5}$ and AOD observations using Pearson correlation coefficients and mean percentage error $MPE = \frac{1}{N}\sum(\frac{O-M}{O})$



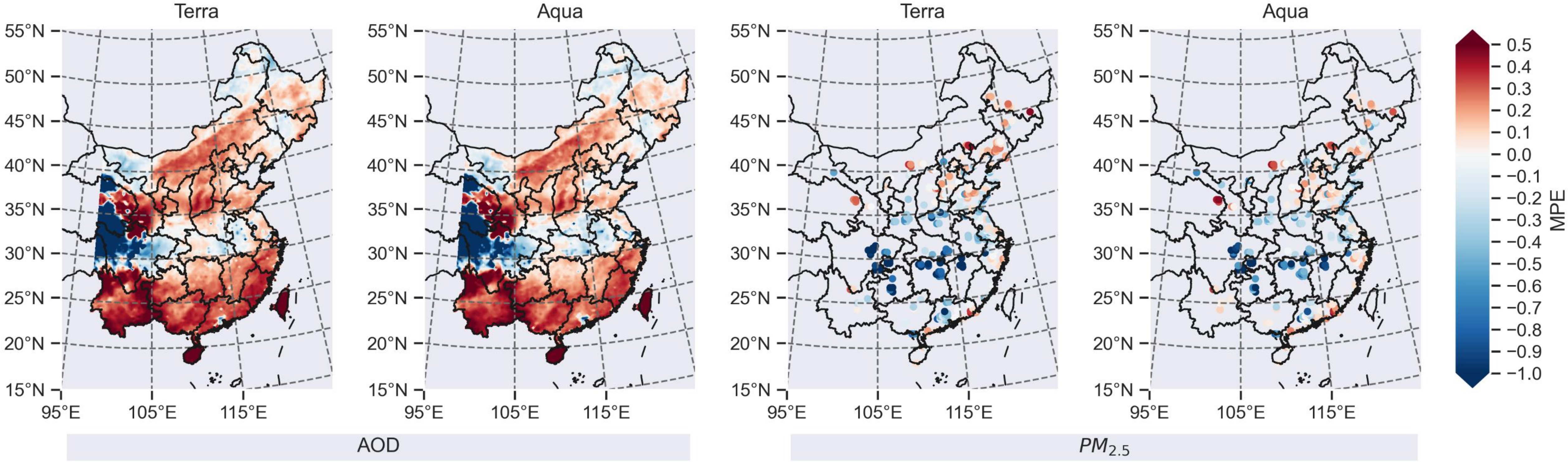
GEOS-Chem model evaluations

Spatial perspective against $PM_{2.5}$ and AOD observations using Pearson correlation coefficients



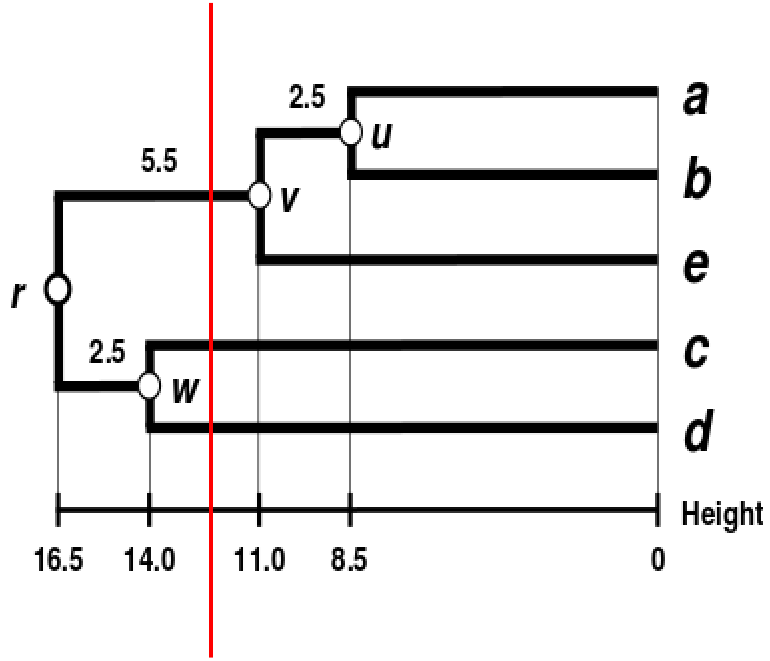
GEOS-Chem model evaluations

Spatial perspective against $PM_{2.5}$ and AOD observations using mean percentage error $MPE = \frac{1}{N}\sum(\frac{O-M}{O})$



Data clustering

- Calculate correlation distance $D_{ij} = \frac{1}{N}\sum_{k=1}^{N}(1 - R_{ijk})$ between $PM_{2.5}$ monitoring locations.
- Merge closest locations and update distance matrix ($D_{A \cup B,X} = \frac{|A|D_{A,X}+|B|D_{B,X}}{|A|+|B|}$) iteratively to form a dendrogram.
- Choose a distance threshold (e.g. 0.5) to obtain clusters and expand to the nearby space.



	a	b	c	d	e
a	0	17	21	31	23
b	17	0	30	34	21
c	21	30	0	28	39
d	31	34	28	0	43
e	23	21	39	43	0

	(a,b)	c	d	e
(a,b)	0	25.5	32.5	22
c	25.5	0	28	39
d	32.5	28	0	43
e	22	39	43	0

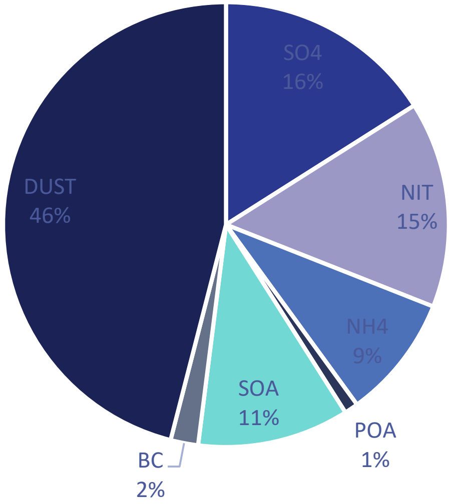
	((a,b),e)	c	d
((a,b),e)	0	30	36
c	30	0	28
d	36	28	0

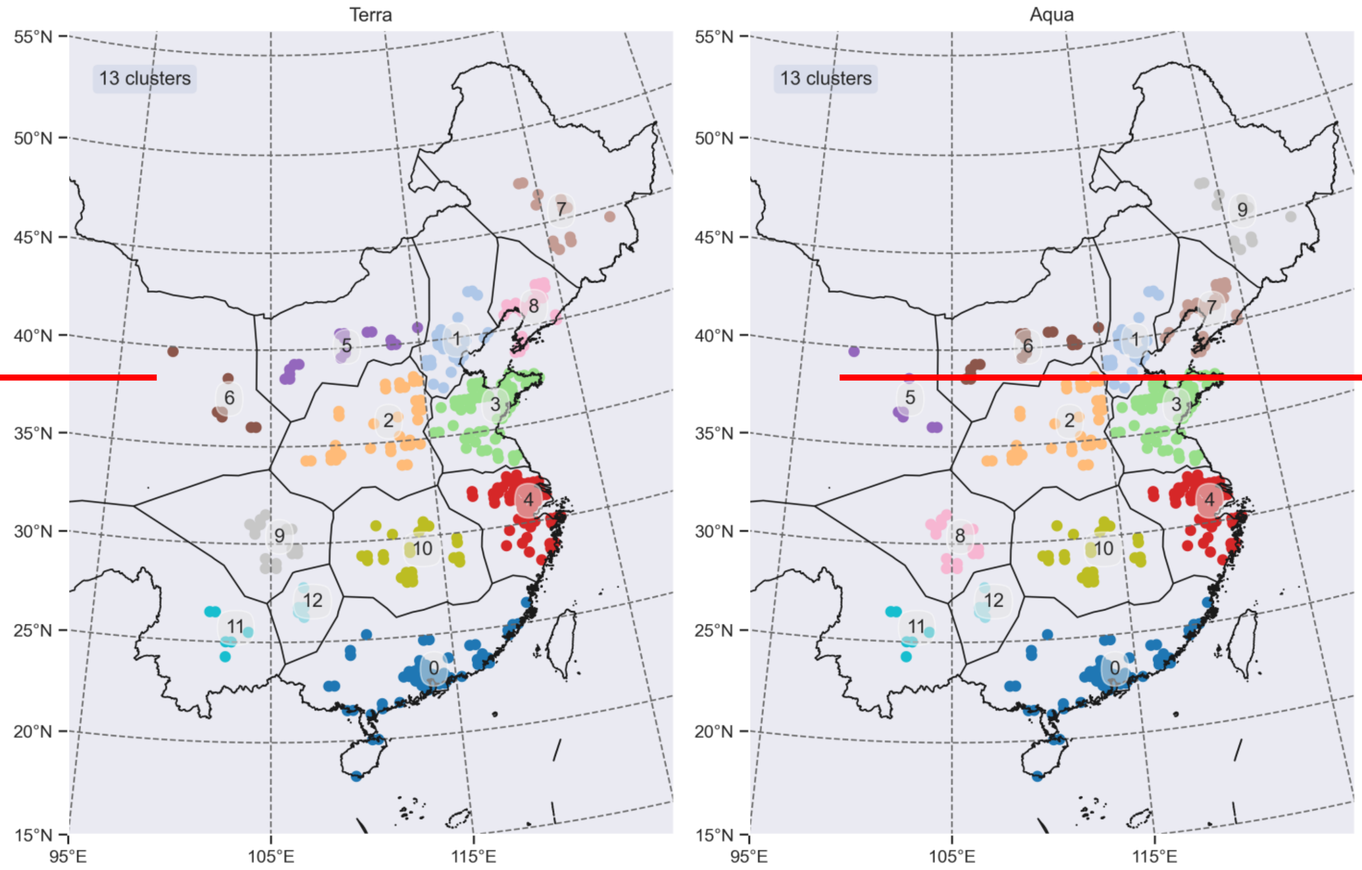
	((a,b),e)	(c,d)
((a,b),e)	0	33
(c,d)	33	0

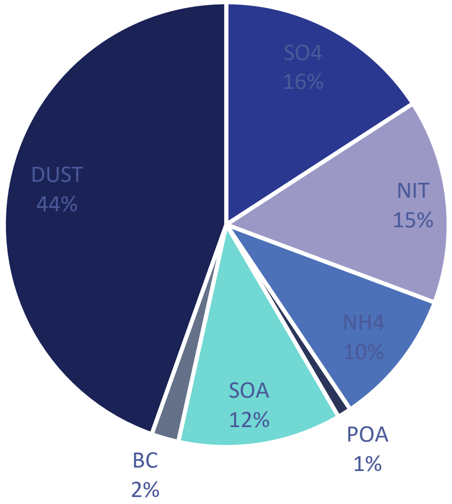
Wikipedia

Data clustering

Annul scale: boundaries similar to those of urban agglomerations. Most but north-western clusters are dominated by secondary PM.

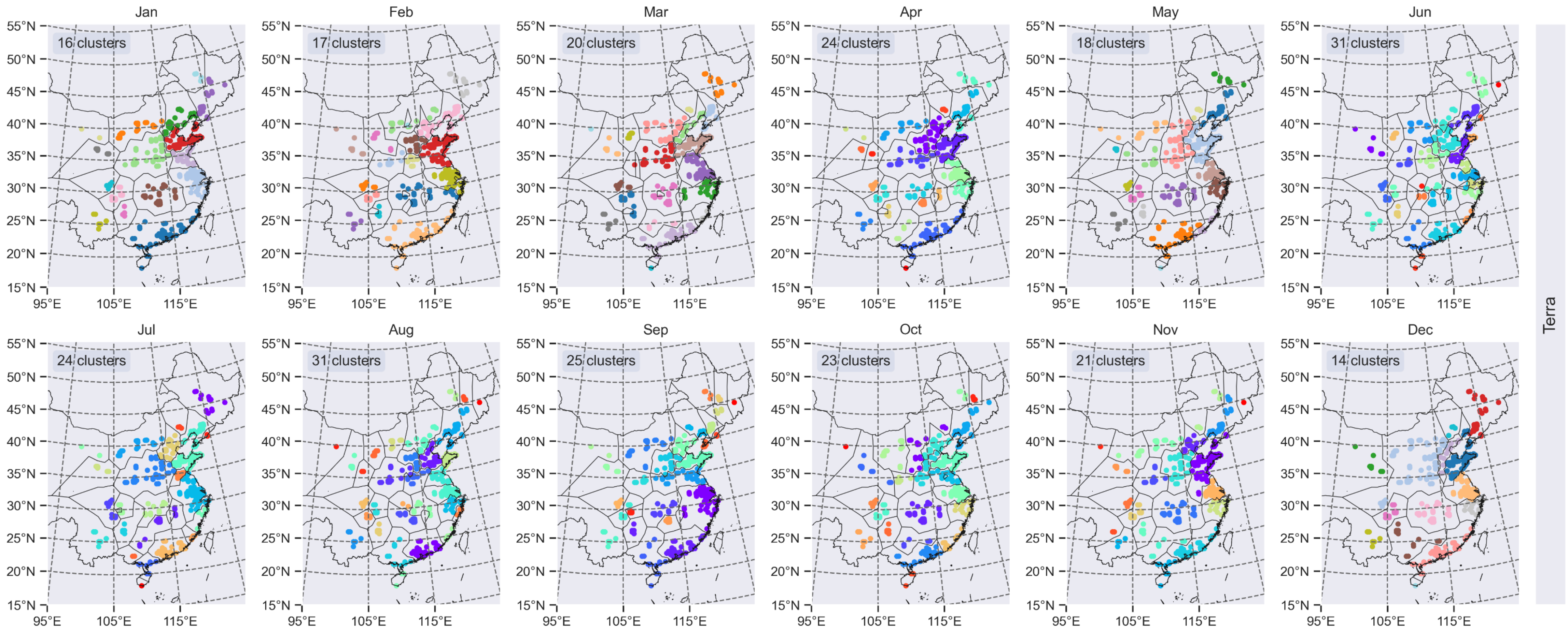






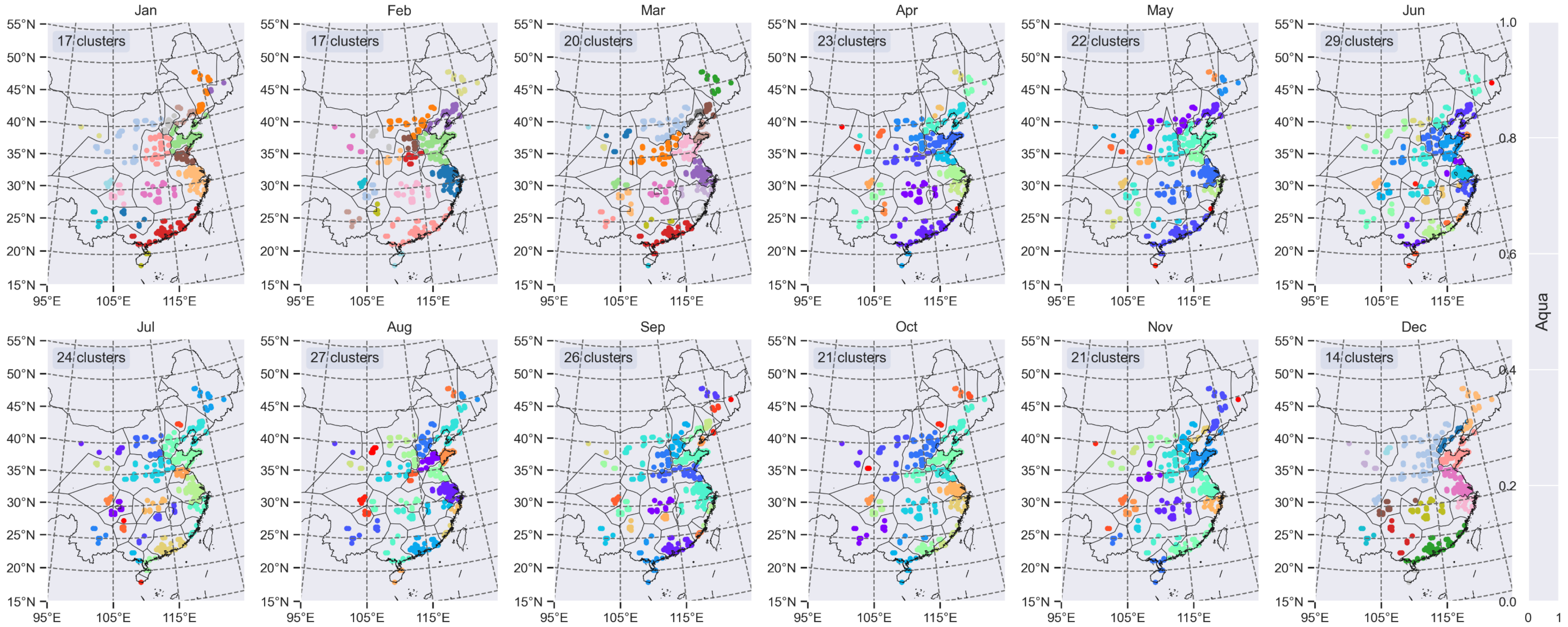
Data clustering

Monthly scale: more fragmented during warm than cold months irrespective of local overpass times.



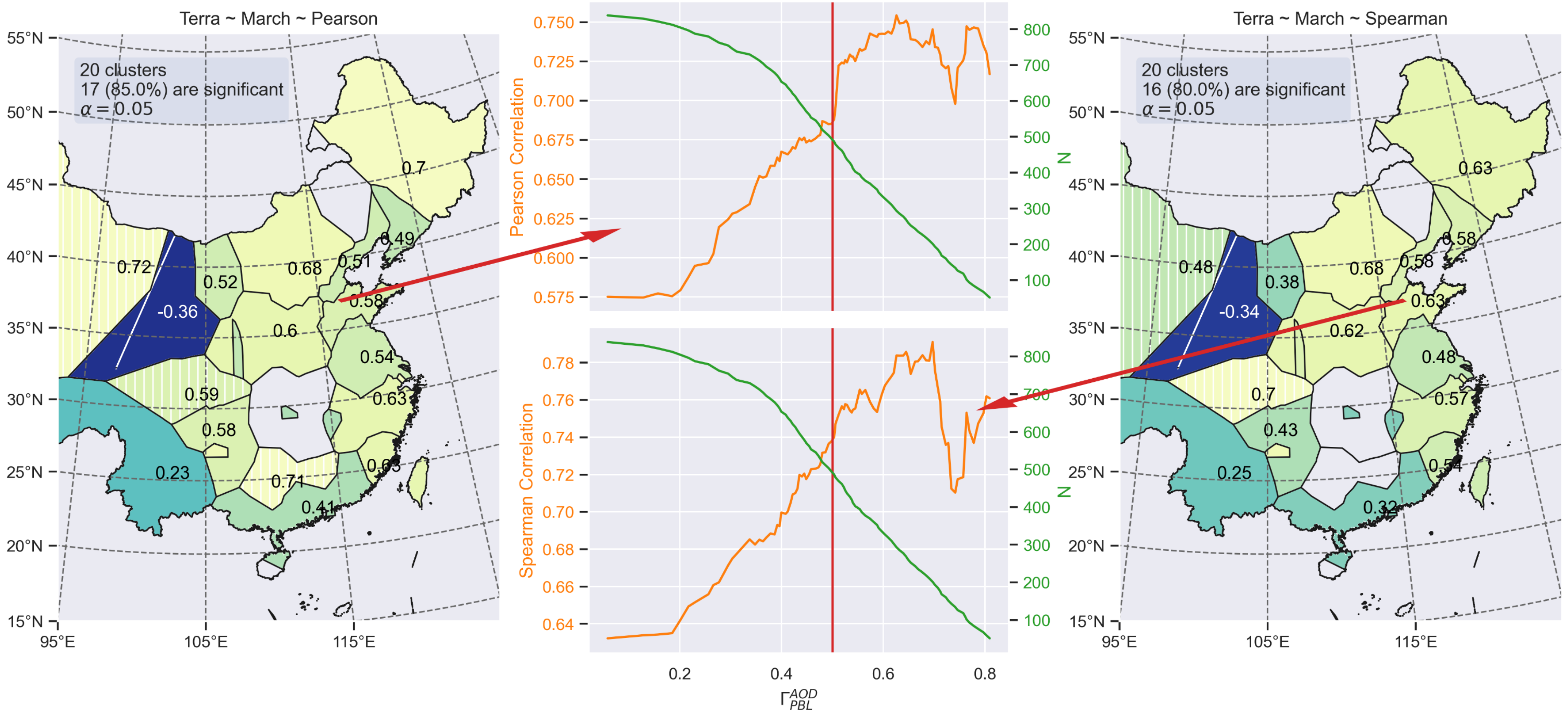
Data clustering

Monthly scale: more fragmented during warm than cold months irrespective of local overpass times.



Data suitability

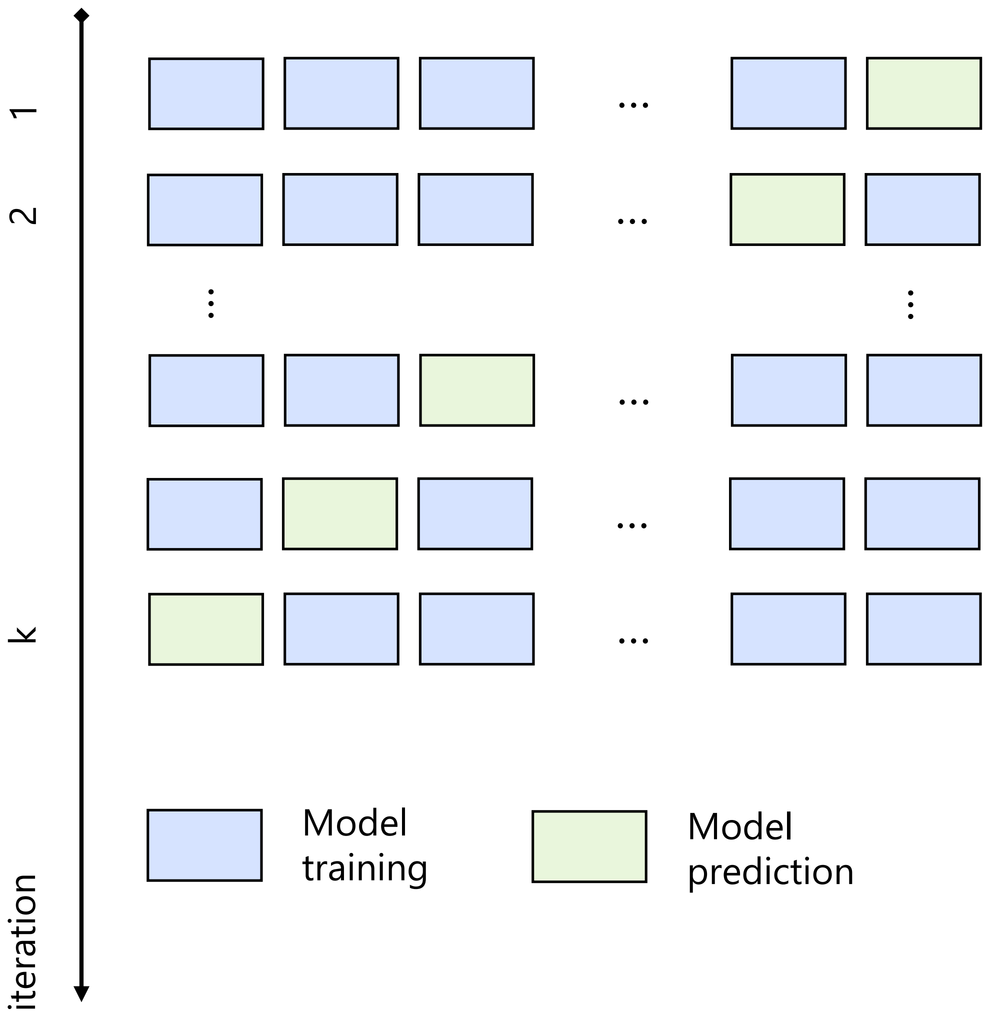
$PM_{2.5}$-AOD correlations generally increases with increasing $\Gamma_{PBL}^{AOD} = \frac{PBL\ AOD}{Column\ AOD}$.



Data-driven model development

- PooledOLS: Pooled Ordinary Least Squares
- TFEM: Time fixed effects model
- RF1 and RF2: Random forest model including day of year or not

$$
\begin{aligned}
PM_{2.5}{}_g^d = \quad & \beta_0 + \beta_{AOD} AOD_g^d + \beta_{PBLH} PBLH_g^d \\
& + \beta_{RH\_PBL} RH\_PBL_g^d + \beta_{TS} TS_g^d \\
& + \beta_{PRECTOT} PRECTOT_g^d + \beta_{U10M} U10M_g^d \\
& + \beta_{V10M} V10M_g^d + \beta_{SLP} SLP_g^d + \epsilon_g^d,
\end{aligned}
$$



Data-driven model development

- The model trained by the suitable data reduces model bias using full data by 10-15% and 9-12% for the Terra and Aqua overpass times, respectively.
- The model trained by the suitable data improves model CV-R^2 using full data by up to 8% and 5% for the Terra and Aqua overpass times, respectively.

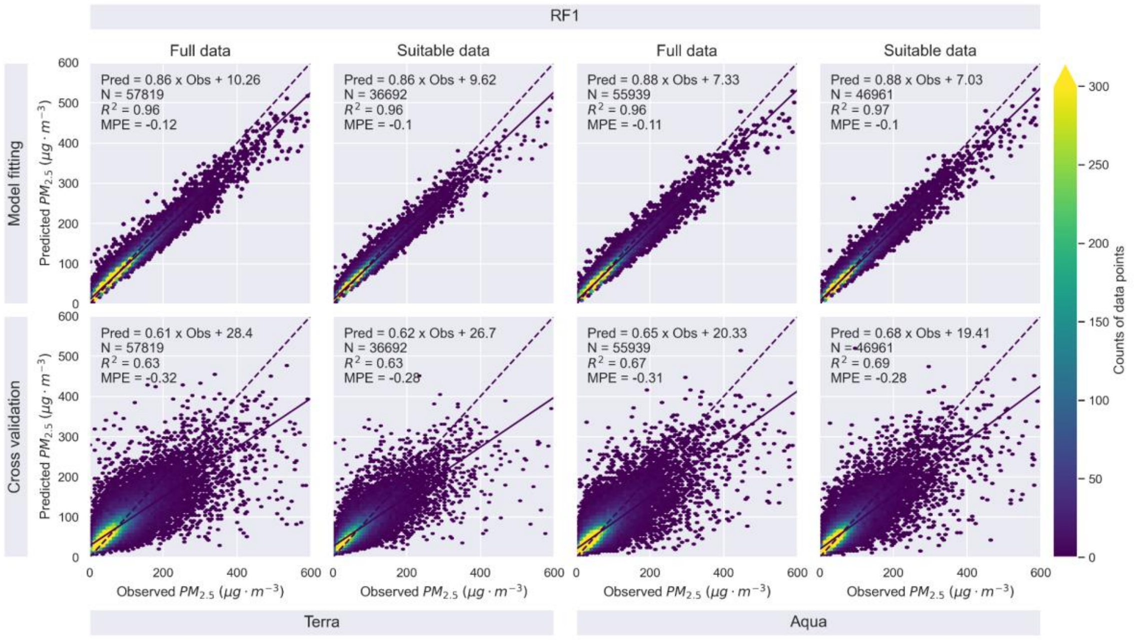
Table 1

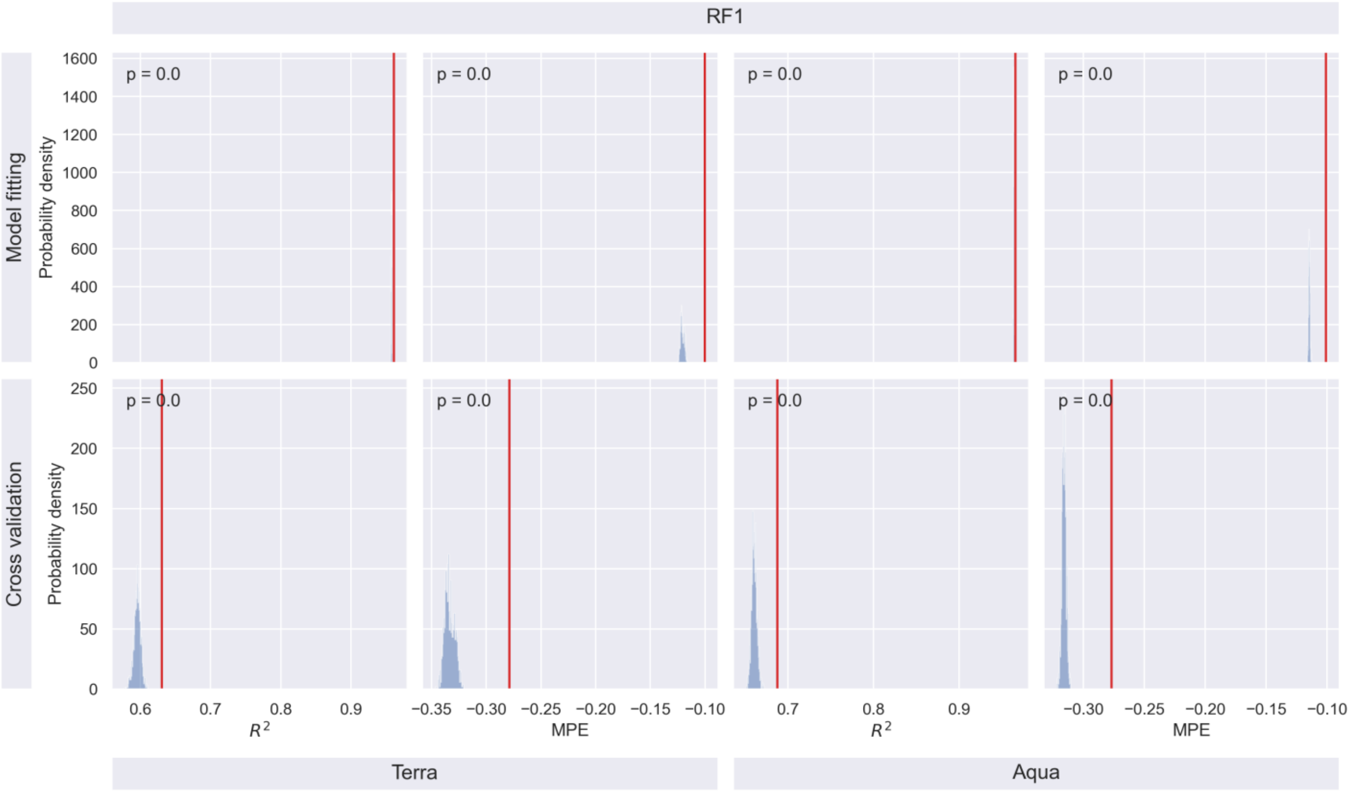
Overall model fitting and cross validation results of the PooledOLS, TFEM, RF1, and RF2 models in 2014 over eastern China. N, R^2, and MPE represent statistics of the model trained by the full data, while N′, $R^{2'}$, and MPE′ represent those of the model trained by the suitable data. R_p^2 and MPE_p represent the possibility of obtaining a model performance no worse than that of our approach by randomly selecting a subset of the full data that matches the length of the suitable table to train the model.

	Model	Terra								Aqua							
		N	N′	R^2	$R^{2'}$	R_p^2	MPE	MPE′	MPE_p	N	N′	R^2	$R^{2'}$	R_p^2	MPE	MPE′	MPE_p
MF	PooledOLS	57819	36692	0.37	0.39	0.0	−0.48	−0.41	0.0	55939	46961	0.43	0.45	0.0	−0.45	−0.41	0.0
	PanelOLS	57819	36692	0.63	0.67	0.0	−0.25	−0.21	0.0	55939	46961	0.69	0.71	0.0	−0.24	−0.21	0.0
	RF1	57819	36692	0.96	0.96	0.0	−0.12	−0.10	0.0	55939	46961	0.96	0.97	0.0	−0.11	−0.10	0.0
	RF2	57819	36692	0.97	0.96	0.0	−0.10	−0.09	0.0	55939	46961	0.97	0.97	0.0	−0.10	−0.09	0.0
CV	PooledOLS	57819	36692	0.36	0.39	0.0	−0.48	−0.41	0.0	55939	46961	0.43	0.45	0.0	−0.45	−0.41	0.0
	PanelOLS	57819	36692	0.58	0.58	0.0	−0.27	−0.24	0.0	55939	46961	0.64	0.66	0.0	−0.26	−0.23	0.0
	RF1	57819	36692	0.63	0.63	0.0	−0.32	−0.28	0.0	55939	46961	0.67	0.69	0.0	−0.31	−0.28	0.0
	RF2	57819	36692	0.68	0.66	0.0	−0.29	−0.26	0.0	55939	46961	0.73	0.73	0.0	−0.28	−0.25	0.0

Data-driven model development

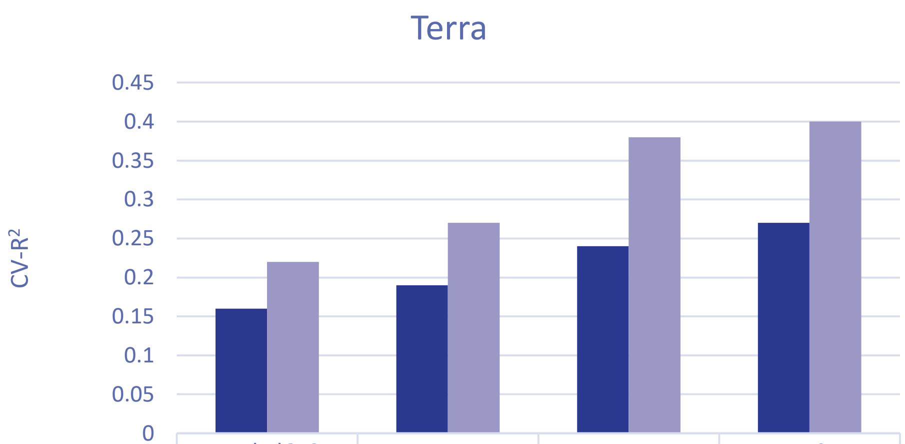
- All these comparative models improvements are statistically significant ($p<0.05$) confirmed by a Monte Carlo simulation.



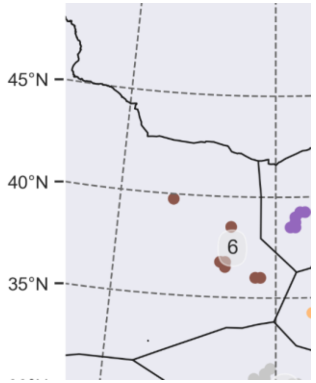


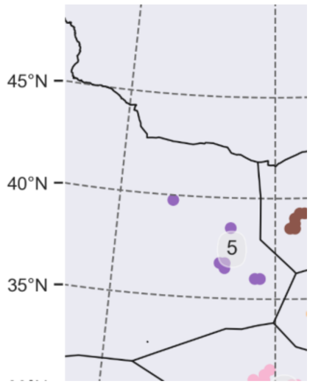
Data-driven model development

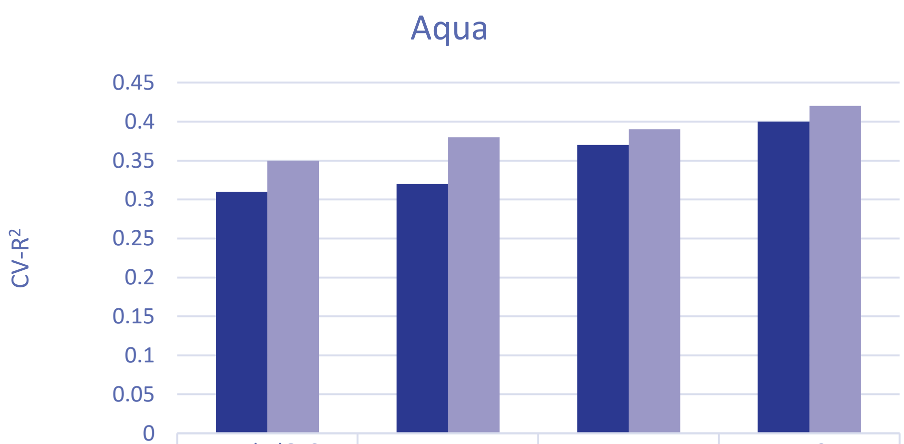
- We see consistent reductions in model bias in almost all clusters. The same goes to CV-R^2 despite some fluctuations during Terra overpass time. Regions dominated by natural aerosols such as dust are particularly distinguishable.



	PooledOLS	TFEM	RF1	RF2
Full data	0.16	0.19	0.24	0.27
Suitable data	0.22	0.27	0.38	0.4





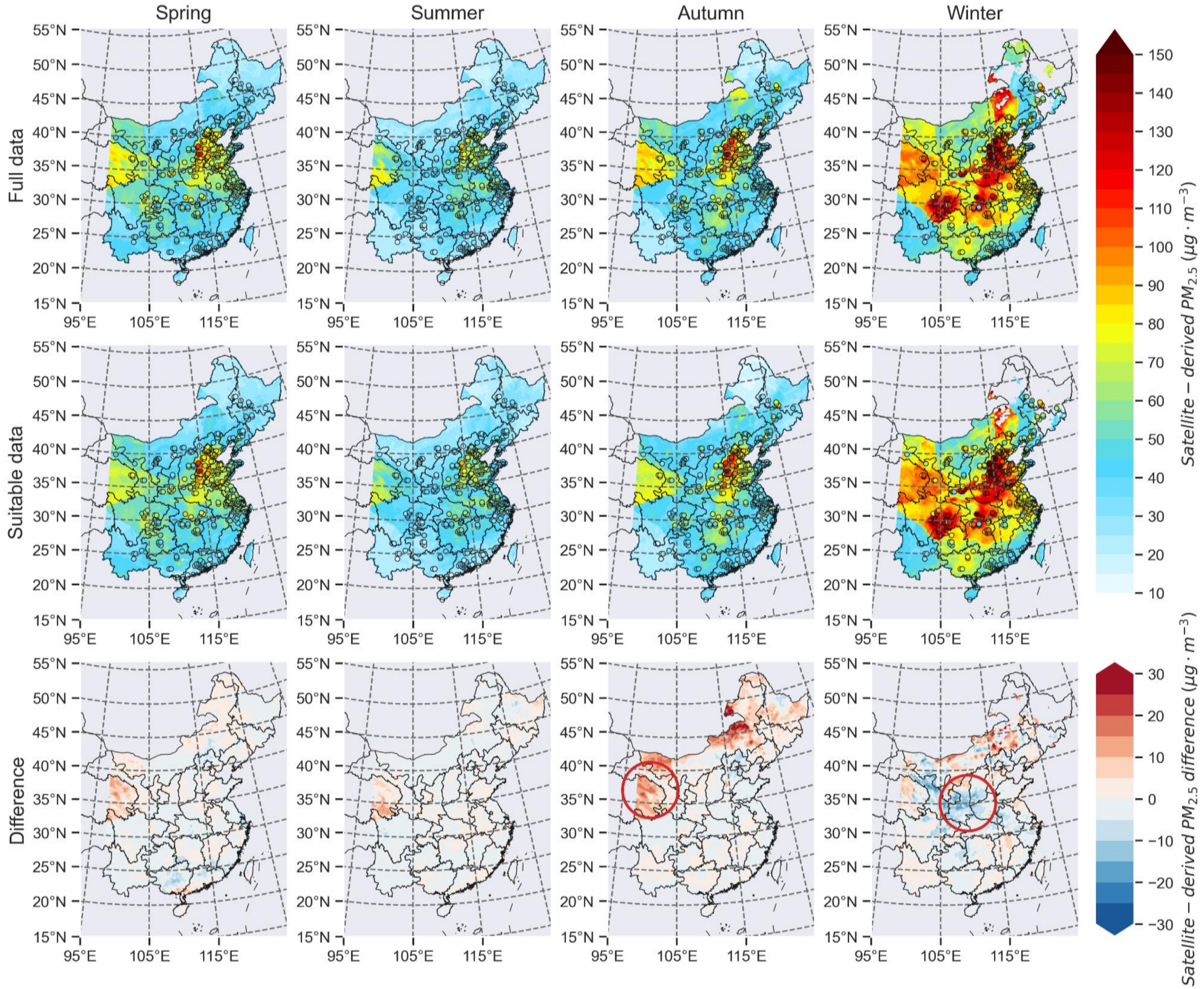


	PooledOLS	TFEM	RF1	RF2
Full data	0.31	0.32	0.37	0.4
Suitable data	0.35	0.38	0.39	0.42

$PM_{2.5}$ mapping

Seasonal $PM_{2.5}$ maps are accordingly improved, with bias of the:

- autumn $PM_{2.5}$ estimates over Qinghai and Gansu provinces reduces from -8% to -5%;
- winter $PM_{2.5}$ estimates over Shaanxi, Shanxi, and Henan provinces reduces from 11% to 6%.



Discussion and concluding remarks

- The proposed model framework that considers the representativeness of AOD for $PM_{2.5}$ reduces bias in $PM_{2.5}$ estimates by 9-15% and captures more variations in $PM_{2.5}$ by up to 8%.
- The resulting $PM_{2.5}$ estimates can be incorporated into a data assimilation system where gaps are filled by a dynamic model of aerosols.
- The proposed model framework is sufficiently generic in that it can be applied to other periods and regions of interest with appropriate process- and data-driven models defined.