Improving representation of the AOD to $PM_{2.5}$ relationship with a convolutional neural network

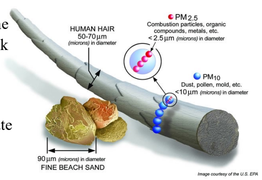
Siyuan Shen[1], Aaron van Donkelaar[1], Chi Li[1], Nathan Jacobs[2], Chenguang Wang[2], and Randall V. Martin[1]

1. Department of Energy, Environmental, and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;

2. Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;

Motivation

Exposure to ambient **fine particulate matter ($PM_{2.5}$)** is the leading global environmental risk factor for mortality and disease burden.



Lack of ground monitors motivate us to get a reliable estimation of global $PM_{2.5}$ concentration.

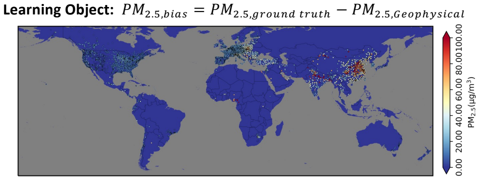
Deep learning is a powerful tool, with growing applications in many fields. Although traditional methods such as geographically weighted regression (GWR) [1] have been proven to be powerful methods for globally representing the residual bias in geophysical satellite-derived $PM_{2.5}$ vs observations, we still seek to improve accuracy through deep learning models.

Methods

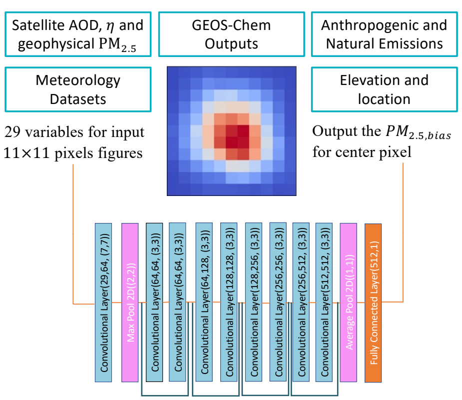
Initial method to estimate global ground-level $PM_{2.5}$, geophysical $PM_{2.5}$ based on satellite derived AOD and GEOS-Chem model[2]. Our model is based on the geophysical $PM_{2.5}$ concentration.

$$PM_{2.5,Geophysical} = \eta \times AOD_{Retrived} \qquad \eta = \frac{PM_{2.5,Model}}{AOD_{Model}}$$

Learning Object: $PM_{2.5,bias} = PM_{2.5,ground\ truth} - PM_{2.5,Geophysical}$

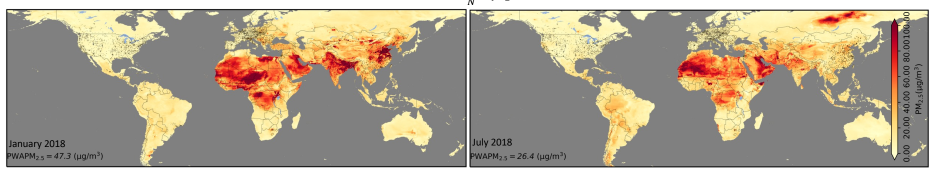


Location of the monitors.



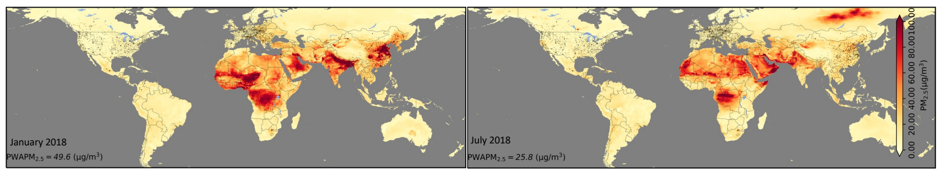
Results – Performance Evaluation

Simple Loss Function – **Mean Square Error(MSE)** Loss Function: $L = \frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2$



Loss Function with a priori value constrains :

$$L = \frac{1}{N}\sum_{i=1}^{N}[(1+\beta e^{-\alpha y_i^2})(f(x_i)-y_i)^2 + \lambda_1 ReLU(-f(x_i) - GeoPM_{2.5,i}) + \lambda_2 ReLU(f(x_i) - \gamma GeoPM_{2.5,i})], \alpha, \beta, \gamma, \lambda_1, \lambda_2 > 0$$



Adjusted Cost functions with a priori value constrains improve the gridded $PM_{2.5}$ estimation in areas with sparse monitors, e.g., Tibet.

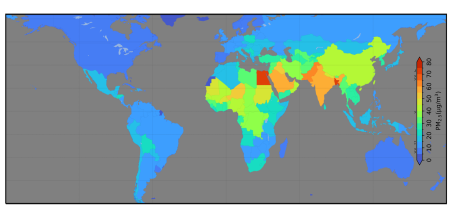
It also improves the coefficient of determination(R^2) in areas with low concentration, e.g., North America, and Europe.

Spatial Cross-Validation results compared with ground observation

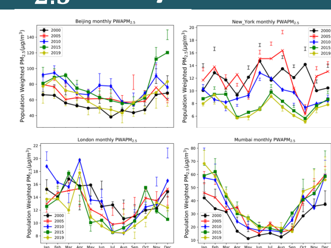
Orange – Global; Green – Asia; Blue – North America; Grey - Europe

	CNN R^2 Global, 2015-2019, N=10870 MSE	CNN R^2 Global, 2015-2019, N=10870 w/Penalties	Hybrid R^2 Global, 2015-2019, N=10870,	Asia, 2015-2019, N=3515, MSE	Asia, 2015-2019, N=3515 w/Penalties	Hybrid R^2 Asia, 2015-2019, N=3515	North America, 2001-2019, N=2874 MSE	North America, 2001-2019, N=2874 w/Penalties	Hybrid R^2 North America, 2001-2019, N=2874	Europe, 2010-2019, N=3310 MSE	Europe, 2010-2019, N=3310 w/Penalties	Hybrid R^2 Europe, 2010-2019, N=3310
Annual	0.86 [0.83, 0.89]	0.86 [0.82, 0.89]	0.84 [0.81,0.86]	0.74 [0.62, 0.78]	0.74 [0.62,0.78]	0.69 [0.65,0.72]	0.58 [0.45,0.70]	0.59 [0.47,0.69]	0.57 [0.42,0.73]	0.69 [0.60,0.80]	0.70 [0.60,0.79]	0.68 [0.66,0.70]
January	0.89 [0.86,0.92]	0.89 [0.86,0.92]	0.86 [0.84,0.88]	0.80 [076,0.84]	0.79 [0.71,0.81]	0.75 [0.72,0.80]	0.55 [0.39,0.64]	0.50 [0.39,0.57]	0.51 [0.29,0.64]	0.74 [0.66,0.85]	0.75 [0.66,0.85]	0.72 [0.67,0.76]
April	0.85 [0.82,0.88]	0.84 [0.81,0.88]	0.81 [0.78,0.84]	0.67 [0.63,0.75]	0.66 [0.60,0.72]	0.62 [0.58,0.66]	0.55 [0.46,0.70]	0.60 [0.49,0.71]	0.57 [0.29,0.64]	0.61 [0.53,0.71]	0.62 [0.54,0.68]	0.59 [0.53,0.64]
July	0.78 [0.74,0.81]	0.77 [0.72,0.81]	0.72 [0.67,0.77]	0.63 [0.55,0.73]	0.64 [0.55,0.73]	0.59 [0.49,0.65]	0.62 [0.52,0.71]	0.66 [0.55,0.74]	0.61 [0.43,0.78]	0.55 [0.45,0.65]	0.57 [0.47,0.67]	0.53 [0.42,0.58]
October	0.83 [0.78,0.87]	0.83 [0.77,0.88]	0.78 [0.75,0.82]	0.71 [0.55,0.77]	0.71 [0.53,0.75]	0.61 [0.57,0.64]	0.53 [0.45,0.67]	0.57 [0.48,0.68]	0.53 [0.42,0.69]	0.69 [0.62,0.80]	0.70 [0.63,0.80]	0.67 [0.62,0.71]

Results – Exposure to $PM_{2.5}$ Analysis

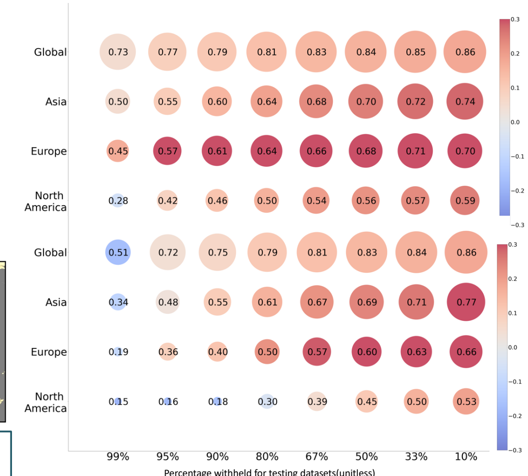


Map of population-weighted annual average $PM_{2.5}$ **in 2018**



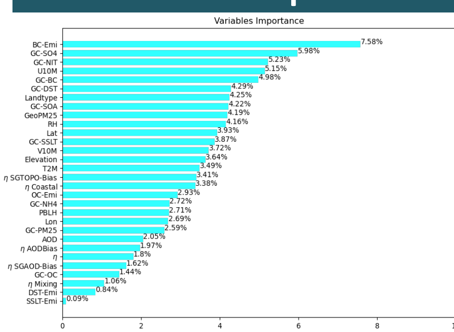
Population-weighted annual average $PM_{2.5}$ seasonal trends in Beijing, New York, London, and Mumbai

Robustness Test



We withheld different percentage of the monitors for testing datasets from 10% to 99%. The model includes a priori estimation of $PM_{2.5}$(above) show more robustness than the model without a priori estimation of $PM_{2.5}$(bottom).

Variables importance



Conclusions:

- Our method shows better performance than both traditional statistical method(GWR), and simple deep learning model.
- Our model shows high accuracy with only few ground monitors which indicates the reliability of the estimation in area with sparse monitors.

Acknowledgement:

NASA Grant 80NSSC22K0200

References

(1) van Donkelaar, Aaron, et al. "Monthly global estimates of fine particulate matter and their uncertainty." *Environmental Science & Technology* 55.22 (2021): 15287-15300.

(2) Van Donkelaar, Aaron, et al. "Estimating ground-level PM2. 5 using aerosol optical depth determined from satellite remote sensing." *Journal of Geophysical Research: Atmospheres* 111, no. D21 (2006).

Contact – Siyuan Shen: s.siyuan@wustl.edu