

Exploring the health disparity of long-term $PM_{2.5}$ exposure with advanced $PM_{2.5}$ mapping and hospital admission records

Ying Liu ¹ and Guofeng Cao ²

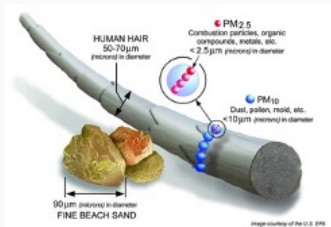
AutoCarto, November 2022

¹Department of Environmental and Occupational Health
University of Montreal
`ying.liu.2@umontreal.edu`

²Department of Geography
University of Colorado Boulder
`guofeng.cao@colorado.edu`
<https://www.spatstat.com>

Fine Particulate Matter ($PM_{2.5}$)

- A complex mixture of extremely small particles and liquid droplets in the air with aerosol dynamic diameters equal to or less than 2.5 micrometers;
- An estimation of 3.1 million premature deaths occurs globally associated with $PM_{2.5}$ exposure every year (Cohen et al., 2017);
- In the U.S., the disparities in exposure to $PM_{2.5}$ and related health outcomes are known to exist across population and income groups (Colmer et al., 2020; Jbaily et al., 2022)



$\text{PM}_{2.5} \text{ morbidity risks} = f(\text{PM}_{2.5} \text{ exposure, morbidity statistics})$

- Typical studies have been performed based on the limited ground-based $\text{PM}_{2.5}$ monitoring observations or coarse resolution $\text{PM}_{2.5}$ images;
- Most of the studies focus on short-term effects of $\text{PM}_{2.5}$ exposure in limited geographic regions (e.g., cities) mainly because long-term disease statistics difficult to obtain for large geographic regions.

The goal of this study

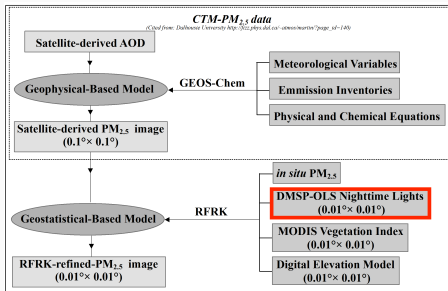
- to create a high resolution mapping of $\text{PM}_{2.5}$ concentration maps;
- to explore the long-term $\text{PM}_{2.5}$ -morbidity and disparities in the state of Arizona with 10 years of hospital admission records

High-resolution mapping of $\text{PM}_{2.5}$

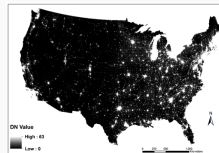
- Statistical models
 - Build statistical relationship between $\text{PM}_{2.5}$ ground-truth measurements, AOD and other relevant variables
 - e.g., land use regression, geographically weighted regression
 - Advantages: easy to implement
 - Limitations: availability of the ground-truth measurements
- Geophysical models
 - Model the diffusion of chemical compositions in atmospheric dynamics
 - e.g., GEOS-Chem chemical transport model to convert the AOD components to $\text{PM}_{2.5}$
 - Advantages: not directly rely on $\text{PM}_{2.5}$ ground-truth measurements
 - Limitations: low spatiotemporal resolution due to the high computational load and incompleteness of input data source (e.g., national emission inventory)

A machine learning-based geostatistical downscaling approach

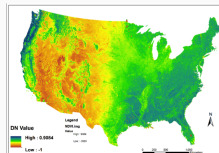
Flowchart of the proposed approach



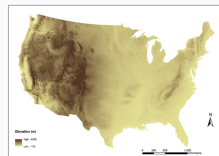
Liu et al. (2018)



NTL



NDVI



Elevation

High-resolution mapping of $PM_{2.5}$

Machine learning-based regression kriging for downscaling:

$$\hat{z}(s)_{RFRK} = \mathcal{F}_{RF}[\mathbf{x}(s); \hat{\beta}] + \epsilon(s),$$

$$\text{where } \epsilon(s) \sim \mathcal{GP}(\mu, \Sigma)$$

- Machine learning (random forests) for non-linear relationship between $PM_{2.5}$ and covariates
- Kriging for complex spatial effects
- Practical two step regression kriging

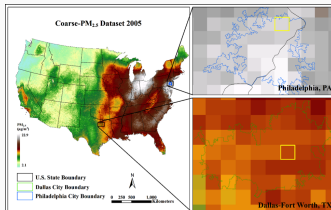
Apply the RFRK to refining the coarse- $PM_{2.5}$ dataset:

$$\hat{z}_{RFRK}^{1km} = \mathcal{F}_{RF}(z_{10km}^{1km}, \mathbf{x}_{ntl}^{1km}, \mathbf{x}_{ndvi}^{1km}, \mathbf{x}_{ele}^{1km}; \hat{\beta}) + \epsilon$$

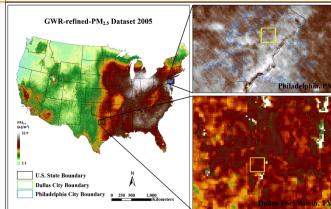
Liu et al. (2018)

High-resolution mapping of $PM_{2.5}$

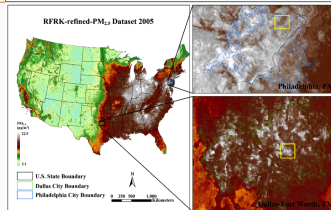
GEOS (10km)



GWR (1km, WUSTL)



RFRK (1km, our results)

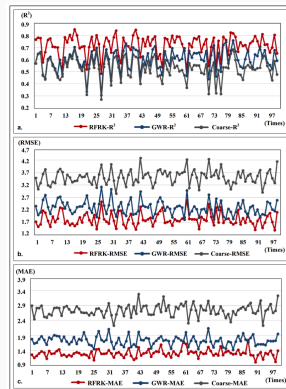


High-resolution mapping of $PM_{2.5}$

Accuracy comparison

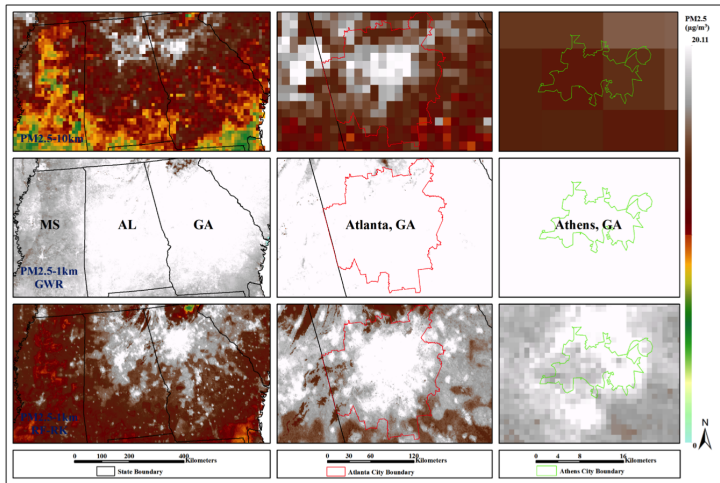
- Compared with a recently refined $PM_{2.5}$ dataset with 1 km resolution using geographically weighted regression (GWR)
- 100 cross-validations with 90% of ground-truth as training and the rest 10% for validation

	R^2	ME ($\mu g/m^3$)	MAE ($\mu g/m^3$)	$RMSE$ ($\mu g/m^3$)	t -test	F -test
Coarse- $PM_{2.5}$	0.533	2.540	2.749	3.522	0.000***	0.000***
GWR-refined- $PM_{2.5}$	0.600	0.215	1.732	2.299	0.000***	0.000***
RFRK	0.720	-0.046	1.309	1.836	/	/

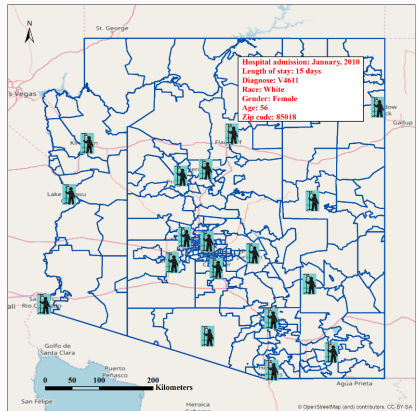
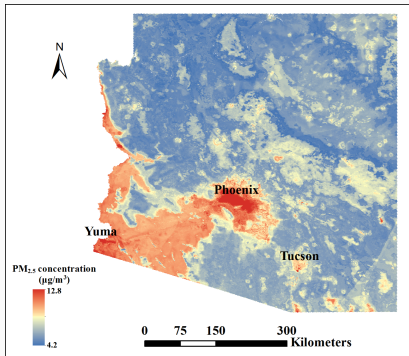


High-resolution mapping of $PM_{2.5}$

Improved spatial variations (GEOS vs. GWR vs. RFRK):



PM_{2.5} maps and hospital admission records for Arizona



- State Inpatient Database (SID) from Healthcare Cost Utilization Project (HCUP)
- Years: 2005-2014
- Record number: 7.7 million
- Zip-code level

SID data

- Within one year, any patients with the same birth month and year, sex, race, married status, economic status, insurance type, town of residence, and zip code of residence are considered as one person.
- The cleaned SID data are classified into groups by sex, racial/ethnicity group, and diagnoses (ICD-9-CM code):
 - Asthma
 - Skin cancer
 - Chronic respiratory diseases
 - Heart diseases
 - Cerebrovascular diseases

PM_{2.5} exposure estimation

$$EX = \frac{PM_p + PM_c \times \frac{m}{12}}{\left(1 + \frac{m}{12}\right)}$$

- PM_p is the average PM_{2.5} concentration of the year prior to the year of a patient's hospitalization
- PM_c is the average PM_{2.5} concentration of the year when the patient visits the hospital
- m denotes the m -th month of the year when the patient visited the hospital.

Population data

- Demographic data are retrieved from the US Census Bureau (USCB) that contain racial and ethnic groups (African American, non-Hispanic White, Hispanic) and sex (male and female) for each zip code area in Arizona.

Disease specific relative risk

$$RR^d(c) = \frac{EP_c^d / (EP_c^d + EN_c^d)}{CP_{c_0}^d / (CP_{c_0}^d + CN_{c_0}^d)}$$

- $RR^d(c)$: PM_{2.5}-Morbidity RR under PM_{2.5} concentration c
- EP_c^d (EN_c^d) : number of persons with positive (negative) outcomes in exposure group
- $CP_{c_0}^d$ ($CN_{c_0}^d$) : number of persons with positive (negative) outcomes in control group
- c_0 : theoretical-minimum-risk PM_{2.5} level

Theoretical-minimum-risk $PM_{2.5}$ concentrations

An iterative way to find c_0 :

$$RR^d(c_0) = \frac{EP_{c_0'}^d / (EP_{c_0'}^d + EN_{c_0'}^d)}{CP_{c_0'}^d / (CP_{c_0'}^d + CN_{c_0'}^d)}, \quad \text{where } c_0' = c_0 + 0.1$$

c_0 for diseases of interests:

All-cause	$3.3 \mu\text{g}/\text{m}^3$
Skin cancer	$3.8 \mu\text{g}/\text{m}^3$
Asthma	$3.7 \mu\text{g}/\text{m}^3$
Heart diseases	$3.5 \mu\text{g}/\text{m}^3$
Respiratory diseases	$3.6 \mu\text{g}/\text{m}^3$
Cerebrovascular diseases	$3.6 \mu\text{g}/\text{m}^3$

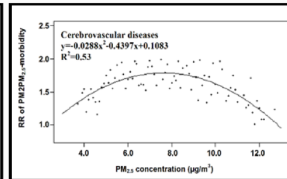
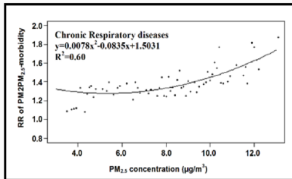
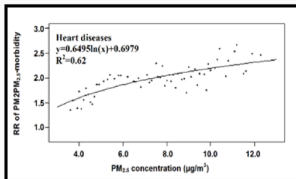
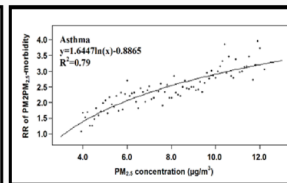
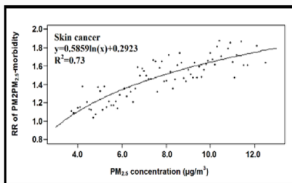
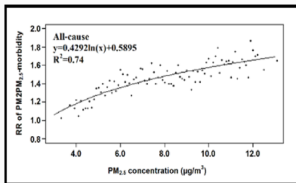
Integrated exposure response

$$RR^d(c) = \begin{cases} 1 + \alpha \left[1 - \exp \left(-\beta (c - c_0)^\delta \right) \right] & c > c_0 \\ 1 & c \leq c_0 \end{cases}$$

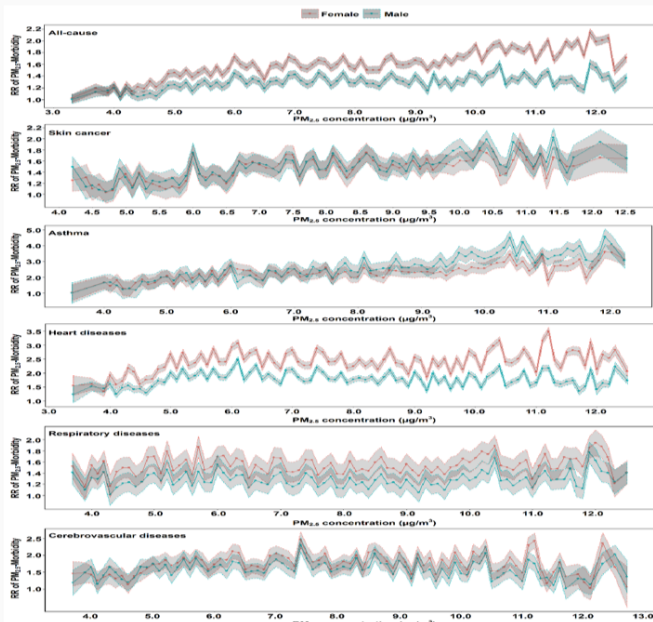
- α , β , and δ are parameters controlling the IER function
- The parameters were provided in GBD 2017 with 95% confidence interval

Burnett et al. (2014)

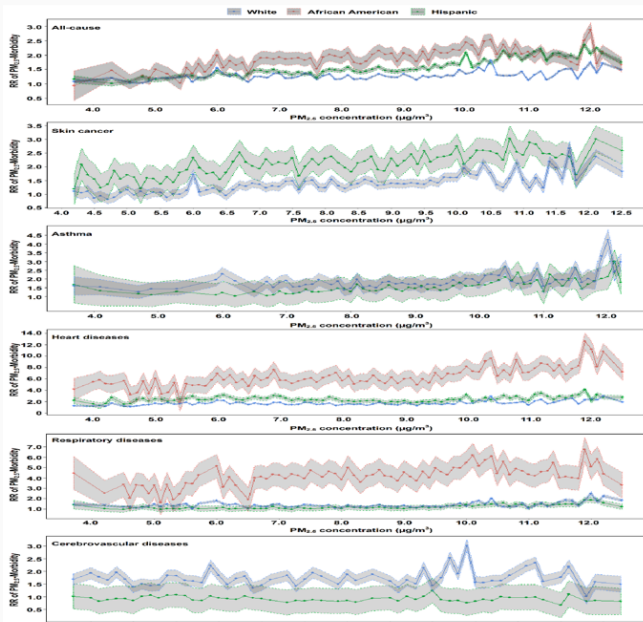
PM_{2.5}-Morbidity RR Functions



RRs of $PM_{2.5}$ -morbidity across Subpopulation Groups



RRs of $PM_{2.5}$ -morbidity across Subpopulation Groups



Summary

- We explored the $PM_{2.5}$ attributable morbidity risk with high resolution mapping of $PM_{2.5}$ concentration and long-term (10 years) of hospital records in the state of Arizona;
- We modeled the $PM_{2.5}$ -morbidity RR for all-cause, skin cancer, asthma, and heart disease are logarithmic and for chronic respiratory and cerebrovascular diseases are polynomial;
- Health disparities of long-term $PM_{2.5}$ exposure across different subpopulation groups:
 - Female is more vulnerable all-cause, heart and respiratory diseases
 - African American has higher risk for all-cause, heart and respiratory disease
 - Hispanic has higher risk skin cancer
 - White has higher risks to cerebrovascular diseases

Limitations and Future Work

- This study only used inpatient data; including more data such as outpatient, emergency room visits can provide comprehensive views;
- The exposure was only derived from ambient $PM_{2.5}$ concentrations, and the method for relative risks was exploratory in nature;
- Uncertainty of data and model was not characterized and analyzed;
- We are working on addressing these issues ...

References

- Burnett, R. T., Arden Pope, C., Ezzati, M., Olives, C., Lim, S. S., Mehta, S., Shin, H. H., Singh, G., Hubbell, B., Brauer, M., Ross Anderson, H., Smith, K. R., Balmes, J. R., Bruce, N. G., Kan, H., Laden, F., Prüss-Ustün, A., Turner, M. C., Gapstur, S. M., Diver, W. R., and Cohen, A. (2014). An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environmental Health Perspectives*, 122(4):397–403.
- Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., Pope, C. A., Shin, H., Straif, K., Shaddick, G., Thomas, M., van Dingenen, R., van Donkelaar, A., Vos, T., Murray, C. J., and Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *The Lancet*, 389(10082):1907–1918.

- Colmer, J., Hardman, I., Shimshack, J., and Voorheis, J. (2020). Disparities in PM_{2.5} air pollution in the United States. *Science*, 369(6503):575–578.
- Jbaily, A., Zhou, X., Liu, J., Lee, T.-H., Kamareddine, L., Verguet, S., and Dominici, F. (2022). Air pollution exposure disparities across US population and income groups. *Nature*, 601(7892):228–233.
- Liu, Y., Cao, G., Zhao, N., Mulligan, K., and Ye, X. (2018). Improve ground-level PM_{2.5} concentration mapping using a random forests-based geostatistical approach. *Environmental Pollution*, 235:272–282.