Improved Estimation of Ambient Air Pollution Exposures: Characterizing Disease Risks and Health Disparities

Jianzhao Bi

Ambient air pollution is a leading health risk factor. In recent decades, increasing wildfire activities and severer heat events have been worsening air pollution across the world, including in the U.S. Accurate exposure estimation is the foundation of reliable assessments of air pollution's impacts on human health. Spatiotemporally resolved exposure data are also essential for identifying disadvantaged communities exposed to disproportionately high pollution. My past and ongoing research has focused on developing improved exposure estimates for air pollution and applying them to accurately characterizing the health risk and disparity of harmful exposures. This line of research is critical as precise and detailed exposures play a central role in conducting reliable epidemiological and environmental justice analyses. My long-term research goal is to extend my analytical knowledge and skills to environmental risk factors with unsatisfactory exposure estimates, including climate-driven air pollutants that are difficult to be characterized (*e.g.*, wildfire smoke) and beyond (*e.g.*, extreme temperatures, noise, and greenness). My future research also aims to investigate how climate-related pollution impacts the health of susceptible (*e.g.*, individuals with or at risk of kidney disease) and disadvantaged populations. I have devoted myself to becoming an interdisciplinary scientist whose work will bridge the fields of exposure assessment and its downstream epidemiological and justice applications.

Research Interests and Contributions

1. Improving air pollution exposure assessment with advanced algorithms and multi-platform data

Ambient air pollution exposure assessment has traditionally relied on regulatory air quality monitoring stations, *e.g.*, the U.S. Environmental Protection Agency (EPA) air quality system (AQS) stations. Due to high instrumentation and maintenance costs, regulatory monitoring is only performed at limited locations (mostly in urban areas) for examining compliance with air quality standards. Given the spatial variability of the pollution, sparse and uneven regulatory monitoring has a limited ability to reflect pollution details, especially in remote communities or when impacted by episodic events such as wildfires. Inaccurate exposure estimates result in non-negligible Berkson- and classical-type errors in epidemiological analysis.

This paradigm is shifting with increasingly available <u>satellite remote sensing instruments</u> to monitor air quality at a global scale. However, there are three issues limiting the application of satellite data to exposure assessment: 1) satellites monitor air pollution over the entire atmospheric column (from the ground to the top of the atmosphere), which is less correlated with ground-level concentrations due to vertical dispersion; 2) satellites usually do not measure the target air pollutants themselves, but their surrogates; for instance, the satellite surrogate of particulate matter (PM), aerosol optical depth (AOD), is only moderately correlated with ground-level PM concentrations; 3) there are substantial missing satellite data (~70% in the U.S.) due to cloud and snow cover. My research has aimed to overcome these challenges to allow the incorporation of satellite data into air pollution exposure estimation. I developed a machine learning model based on the Random Forest algorithm to reliably estimate missing satellite AOD data resulting from cloud and snow cover.[1] The model had satisfactory performance and generated spatiotemporally complete AOD estimates in New York State where ~90% of the data were missing. I further proposed machine learning models to accurately convert AOD to ground-level $PM_{2.5}$ (PM < 2.5 μm in aerodynamic diameters) at a 1-km spatial resolution.[1-3] My proposed methods for satellite AOD have been widely applied to generating high-resolution $PM_{2.5}$ exposures not only in North America, but also in resource-restricted regions in South America, Asia, and Africa.[2, 4-7] My methods have also been applied to forecast near-term air pollution concentrations to support early warning of pollution episodes.[8]

Recently emerged <u>low-cost air sensors</u> (typically cost < $2500) are another promising supplement to regulatory air monitoring due to their lower cost, flexibility of deployment, and ease of maintenance. The development of citizen science has resulted in a large number of low-cost sensors being deployed and maintained by the public with numerous voluntarily collected air quality data. The PurpleAir network with more than 10,000 sensors worldwide is a good example. However, low-cost air quality measurements have larger biases and uncertainties compared to the "gold-standard" regulatory measurements. It is difficult to calibrate the sensors individually in a publicly maintained network such as PurpleAir. My research, for the first time, proposed reliable statistical methods to 1) simultaneously calibrate a large number of low-cost sensors across a large domain to reduce measurement biases and 2) further reduce the negative influences of the sensors' remaining measurement uncertainties (mostly random errors; difficult to be removed by statistical calibration) during exposure prediction.[9-13] This set of methods allows reliable incorporation of dense low-cost sensor data, with considerably more detailed pollution information, into air pollution prediction. My low-cost sensor work particularly holds promise to advance exposure estimation in resource-restricted environments not covered by dense regulatory monitoring. My work also holds promise to characterize pollution sources that have been poorly covered by regulatory monitoring, *e.g.*, wildfire smoke, since low-cost sensors are usually the only available ground monitors near the fire sources.

Satellite remote sensing and low-cost monitoring data have opened up numerous new possibilities for air pollution exposure assessment. My approaches have improved the accuracy of air pollution exposure assessment with the incorporation of these multi-platform data. My ongoing work has been further expanding the applications of these novel data to other pollutants, such as carbon monoxide,[14] ozone,[15] oxides of nitrogen, and ultrafine particles. Increasing epidemiological and environmental justice applications have been benefiting from these improved exposure estimates.[16-18]

2. Assessing impacts of air pollution exposures on various disease outcomes with rich health data

My previous and ongoing epidemiological analyses mostly focused on assessing the acute health impacts of short-term air pollution exposures. Existing acute health analyses have mostly focused on respiratory and cardiovascular diseases. However, there is emerging toxicological evidence showing that inhaled pollutants may affect more distant organs. For example, inhaled nanoparticles have been detected in urine within minutes after exposure, demonstrating that these particles may pass through kidneys and directly affect kidney function. The corresponding epidemiological evidence, however, is limited. On the other hand, $PM_{2.5}$, a major criteria air pollutant, is a complex mixture of chemical species. While the toxicity of total $PM_{2.5}$ has been well investigated, its chemical components remain underexplored. The improved air pollution exposures and rich health data sources originally not collected for research purposes, *e.g.*, electronic health records and hospital billing data, have played an increasingly important role in epidemiological analysis to fill in these critical knowledge gaps.

Based on emergency department (ED) visits and hospitalization records and spatiotemporally high-resolution exposure data in Atlanta, my research reported positive associations between short-term exposure to criteria air pollutants and acute incidence and exacerbation of <u>kidney disease outcomes</u> (acute kidney injury in particular).[19] This analysis was the first to assess the acute impacts of air pollution on kidney disease risks in the U.S. This study also added to the growing epidemiological evidence that fine particles may impact distant organs over the short term. Additionally, I also investigated how <u>overall toxicity of total $PM_{2.5}$ mixture</u> is associated with its chemical composition. With ED visits and hospitalization data in Los Angeles over a 12-y period, my study found that emissions control programs targeting certain air pollution sources (fossil fuel combustion in particular) altered $PM_{2.5}$ composition (with a decreased proportion of sulfate and an increased proportion of organic carbon), which was associated with an increased risk of cardiovascular diseases and a decreased risk of asthma-related outcomes.[20] While previous studies reported that $PM_{2.5}$-cardiorespiratory disease associations may vary by region and population, my study provided new evidence that these associations may also vary by time, and changes in $PM_{2.5}$ composition could be an important driving factor.

My current studies have taken a step further by assessing the impacts of criteria air pollutants on larger populations based on multi-state exposure and ED visits data. One of my recent studies reported positive associations between exposures to major air pollutants ($PM_{2.5}$, $PM_{10-2.5}$, ozone, CO, NO_x/NO_2, and SO_2) and acute asthmatic effects among the U.S. population from 10 states (currently under review in *Environmental Health Perspectives*). This study also found higher impacts of air pollution exposure on children and the elderly. With advanced exposure and rich health data, my ongoing efforts are to further investigate potential health disparities of air pollution exposures associated with sex, race/ethnicity, and urban/rural environments, as well as to identify populations susceptible to climate-driven air pollution (*e.g.*, wildfire smoke).

In general, based on improved exposure estimates and rich health data, my epidemiological research has been filling in critical knowledge gaps regarding different human body systems, the toxicity of an important air pollutant mixture, and potential susceptibility and health disparities among subpopulations.

Future Research

I have been devoting myself to developing innovative and rigorous analytical approaches to advancing the assessment of environmental exposures that are not well characterized. With the advanced environmental exposure data, I plan to conduct in-depth epidemiological and environmental justice analyses that are previously impractical. My long-term research goals include:

1) Developing advanced quantitative methods to accurately characterize the exposures to climate-driven (*e.g.*, wildfire smoke, extreme heat) and built environment (*e.g.*, greenness, noise) risk factors
2) Filling in the gaps of exposure estimates in resource-restricted environments across the world based upon multi-platform data (*e.g.*, satellites, low-cost sensors, and mobile platforms) with advanced statistical methods
3) Assessing acute and chronic impacts of short- and long-term environmental exposures on various diseases and human body systems that are understudied at regional, national, and global scales
4) Investigating inter- and intra-community exposure and health disparities associated with personal (*e.g.*, race/ethnicity, sex), socioeconomic (*e.g.*, income, education), and environmental (*e.g.*, urban/rural) factors

One of my special research interests is wildfire smoke as wildfire activities have significantly increased worldwide, associated with severer and more frequent heavy smoke events. There are no ground air monitoring systems specifically for wildfire smoke, and we have little undertraining of its health impacts compared to other air pollutants. I have been developing advanced statistical approaches to separating wildfire-derived pollution from total air pollution and exploring appropriate epidemiological methods to assess its adverse effects on various disease outcomes (*e.g.*, cardiorespiratory, birth, and mental health outcomes) and on vulnerable populations (*e.g.*, people with or at risk of chronic kidney disease). I have also been applying the improved wildfire smoke predictions to identifying disadvantaged and/or remote communities affected by extreme smoke episodes (*e.g.*, Methow Valley in Washington), to inform and guide early warning and resource allocation and mitigate the smoke impacts on the residents. I have been collaborating with multi-institutional, multidisciplinary researchers and community leaders to develop several intramural and extramural grant proposals on this topic and will develop more in the next few years.

My long-term career objective is to become an environmental health scientist, with interdisciplinary knowledge and experience in environmental exposure, epidemiology, and environmental justice, to bridge exposure science and its downstream applications by exploring optimal approaches to maximizing the use of advanced exposures in environmental health analysis.

References (representative first-author and co-authored publications)

1. **Bi, J.**; Belle, J. H.; Wang, Y., et al., Impacts of snow and cloud covers on satellite-derived $PM_{2.5}$ levels. *Remote Sensing of Environment* 2019, *221*, 665-674.
2. Stowell, J. D.; **Bi, J.**; Al-Hamdan, M. Z., et al., Estimating $PM_{2.5}$ in Southern California using satellite data: factors that affect model performance. *Environmental Research Letters* 2020, *15*, (9), 094004.
3. Huang, K.; **Bi, J.**; Meng, X., et al., Estimating daily $PM_{2.5}$ concentrations in New York City at the neighborhood-scale: Implications for integrating non-regulatory measurements. *Science of the Total Environment* 2019, *697*, 134094.
4. Vu, B. N.; **Bi, J.**; Wang, W., et al., Application of geostationary satellite and high-resolution meteorology data in estimating hourly $PM_{2.5}$ levels during the Camp Fire episode in California. *Remote Sensing of Environment* 2022, *271*, 112890.
5. Wang, L.; **Bi, J.**; Meng, X., et al., Satellite-based assessment of the long-term efficacy of $PM_{2.5}$ pollution control policies across the Taiwan Strait. *Remote Sensing of Environment* 2020, *251*, 112067.
6. Wang, W.; Liu, X.; **Bi, J.**, et al., A machine learning model to estimate ground-level ozone concentrations in California using TROPOMI data and high-resolution meteorology. *Environment International* 2022, *158*, 106917.
7. Zhang, D.; Du, L.; Wang, W.; Zhu, Q.; **Bi, J.**, et al., A machine learning model to estimate ambient $PM_{2.5}$ concentrations in industrialized highveld region of South Africa. *Remote Sensing of Environment* 2021, *266*, 112713.
8. **Bi, J.**; Knowland, K. E.; Keller, C. A., et al., Combining Machine Learning and Numerical Simulation for High-Resolution $PM_{2.5}$ Concentration Forecast. *Environmental Science & Technology* 2022, *56*, (3), 1544-1556.
9. **Bi, J.**; Stowell, J.; Seto, E. Y. W., et al., Contribution of low-cost sensor measurements to the prediction of $PM_{2.5}$ levels: A case study in Imperial County, California, USA. *Environmental Research* 2020, *180*, 108810.
10. **Bi, J.**; Wildani, A.; Chang, H. H., et al., Incorporating Low-Cost Sensor Measurements into High-Resolution $PM_{2.5}$ Modeling at a Large Spatial Scale. *Environmental Science & Technology* 2020, *54*, (4), 2152-2162.
11. **Bi, J.**; Wallace, L. A.; Sarnat, J. A., et al., Characterizing outdoor infiltration and indoor contribution of $PM_{2.5}$ with citizen-based low-cost monitoring data. *Environmental Pollution* 2021, *276*, 116763.
12. **Bi, J.**; Carmona, N.; Blanco, M. N., et al., Publicly available low-cost sensor measurements for $PM_{2.5}$ exposure modeling: Guidance for monitor deployment and data selection. *Environment International* 2022, *158*, 106897.
13. Wallace, L.; **Bi, J.**; Ott, W. R., et al., Calibration of low-cost PurpleAir outdoor monitors using an improved method of calculating $PM_{2.5}$. *Atmospheric Environment* 2021, *256*, 118432.
14. **Bi, J.**; Zuidema, C.; Clausen, D., et al., Within-City Variation in Ambient Carbon Monoxide Concentrations: Leveraging Low-Cost Monitors in a Spatiotemporal Modeling Framework. *Environ Health Perspect* 2022, *130*, (9), 97008.
15. Zhu, Q.; **Bi, J.**; Liu, X., et al., Satellite-based long-term spatiotemporal patterns of surface ozone concentrations in China: 2005-2019. *Environmental Health Perspectives* 2022, *130*, (2), 027004.
16. Gladson, L.; Garcia, N.; **Bi, J.**, et al., Evaluating the Utility of High-Resolution Spatiotemporal Air Pollution Data in Estimating Local $PM_{2.5}$ Exposures in California from 2015-2018. *Atmosphere* 2022, *13*, (1), 85.
17. He, M. Z.; Do, V.; Liu, S.; Kinney, P.; Fiore, A.; Jin, X.; DeFelice, N.; **Bi, J.**, et al., Short-term $PM_{2.5}$ and cardiovascular admissions in NY State: assessing sensitivity to exposure model choice. *Environmental Health* 2021, *20*, (1), 93.
18. Jin, X.; Fiore, A. M.; Civerolo, K.; **Bi, J.**, et al., Comparison of multiple $PM_{2.5}$ exposure products for estimating health benefits of emission controls over New York State, USA. *Environmental Research Letters* 2019, *14*, (8), 084023.

19. **Bi, J.**; Barry, V.; Weil, E. J., et al., Short-term exposure to fine particulate air pollution and emergency department visits for kidney diseases in the Atlanta metropolitan area. *Environmental Epidemiology* 2021, *5*, (4), e164.
20. **Bi, J.**; D'Souza, R. R.; Rich, D. Q., et al., Temporal changes in short-term associations between cardiorespiratory emergency department visits and $PM_{2.5}$ in Los Angeles, 2005 to 2016. *Environmental Research* 2020, *190*, 109967.