# Improving Global-Scale High-Resolution Air Pollution Prediction With Satellite Data

Yang Liu[1], Jianzhao Bi[1], Avani Wildani[1], and Shrey Gupta[1]

[1]*Emory University*

**We propose to develop a decision support tool with a global-scale, modular and flexible Bayesian Hierarchical modeling system at its core to unite different air pollution prediction approaches. This decision tool will provide high-resolution $PM_{2.5}$ predictions worldwide with rigorous error estimates, and be able to adjust to emerging satellite and ground observations, and be transferred to World Health Organization and Institute for Health Metrics and Evaluation for continued operation in their decision support systems.**

The Global Burden of Disease project (GBD) and the World Health Organization (WHO) estimate that exposure to fine particulate matter ($PM_{2.5}$) contributes to millions of premature deaths globally each year. Ambient $PM_{2.5}$ pollution is the $4^{th}$ highest ranking risk factor in East Asia, the $6^{th}$ in South Asia, and the $7^{th}$ in Africa and the Middle East. However, lack of exposure estimates has been a serious limiting factor to evaluate its associated health outcomes in the developing world with little or no long-term surface monitoring. To isolate the $PM_{2.5}$ health effects among various potential confounders, epidemiologists often need to follow a large cohort of people for many years. Even if extensive networks can be established in the near future, relying solely on their observations will likely delay the onset of prospective health effects studies for years.

Since the launch of NASA's Terra satellite in late 1999, satellite-retrieved particle abundance information such as aerosol optical depth (AOD) has emerged as a promising solution to provide worldwide $PM_{2.5}$ exposure estimates. The $PM_{2.5}$–AOD relationship is a function of particle size distribution, composition, and vertical distribution. Two approaches have been proposed to add these constraints in the modeling process. The first is the statistical modeling approach involving the development of empirical statistical models to estimate $PM_{2.5}$ concentrations with AOD, meteorology, and land-use variables [1]. The second approach, known as the scaling models, relies on the $PM_{2.5}$/AOD ratios simulated by chemical transport models (CTMs), then uses satellite AOD to calibrate the simulated $PM_{2.5}$ concentrations [2]. The statistical models are calibrated by ground truth so their accuracy is higher than the scaling models in regions with extensive ground data, while the scaling models provide a straightforward physical basis for relating AOD to surface $PM_{2.5}$ in regions with sparse or no ground data. These two approaches can be united in a Bayesian Hierarchical modeling framework to provide global-scale high-resolution ($\sim$1-km) $PM_{2.5}$ prediction with rigorous error estimates.

## Project Goals and Phases

- *Phase 1*: Due in Spring 2021 – Compiling global air pollution, satellite, and auxiliary (meteorology and land-use) data for model development. Validating the Bayesian Hierarchical model in data-rich modeling domain of North America (Case #1 in Figure 1).
- *Phase 2*: Due in Winter 2021 – Expanding the regional module to other regions (Case #2 in Figure 1) building on knowledge gained in Case #1. To account for the limited ground monitoring, information from the data-rich region will be borrowed through the use of informative prior distributions for the model parameters.
- *Phase 3*: Due in Summer 2022 – Developing a global module by combining continental modules (Case #3 in Figure 1). Predictions between continental boundaries will be driven by both satellite AOD and CTM scaled $PM_{2.5}$ concentrations with calibration parameters obtained by spatially-interpolating those from neighboring regions.

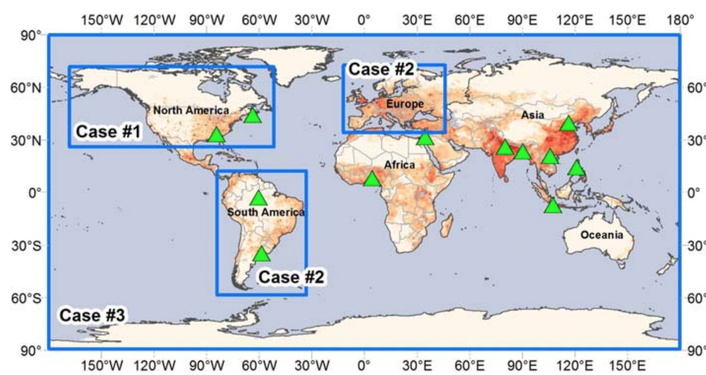


Figure 1: Modeling domains for Cases #1 to #3 (blue boxes).

## Experimentation Requirements: Google Compute Engine

Our experiments deal with processing and training multi-level statistical models with large volumes of data, therefore requiring powerful computation resources such as the Google Compute Engines, which can handle CPU intensive jobs with high demand of memory and disk I/O.

In summary, our project will be the first to develop a global-scale high-resolution air pollution decision support tool based on satellite data. The high-resolution $PM_{2.5}$ predictions can promote implementation of air quality guidelines, environmental health policy, and regulations for human welfare worldwide.

[1] Liu, Y., C. J. Paciorek, and P. Koutrakis (2009), Estimating Regional Spatial and Temporal Variability of $PM_{2.5}$ Concentrations Using Satellite Data, Meteorology, and Land Use Information, *Environ. Health Perspect.*, 117(6), 886-892.

[2] van Donkelaar, A., R. V. Martin, M. Brauer, R. Kahn, R. Levy, C. Verduzco, and P. J. Villeneuve (2010), Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application, *Environ. Health Perspect.*, 118(10), 847-855.