Improving Global-Scale High-Resolution Air Pollution Prediction with Satellite Data - Extension to Carbon Monoxide

Jianzhao Bi[1], Shrey Gupta[2], Avani Wildani[2], and Yang Liu[2]

[1]*University of Washington, Seattle, WA*
[2]*Emory University, Atlanta, GA*

We propose to develop a decision support tool with a global-scale, modular, and flexible statistical modeling system at its core to unite different air pollution prediction approaches. This decision tool will provide high-resolution carbon monoxide (CO) 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 ambient air pollution contributes to millions of premature deaths globally each year [1]. However, lack of exposure estimates for certain air pollutants, especially CO, has been a serious limiting factor to evaluate its associated health outcomes in the developing world with little or no long-term ground-level monitoring. To isolate the CO 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 ground observations will likely delay the onset of prospective health effects studies for years.

Satellite-retrieved air pollution abundance information, such as CO column densities, has emerged as a promising solution to provide worldwide CO exposure estimates. The ground-satellite CO concentration relationship is a function of meteorological factors and CO vertical distribution. A widely adopted approach to add these constraints to the modeling process relies on the ground-satellite concentration ratios simulated by chemical transport models (CTMs) [2-3]. However, based on our existing experience in $PM_{2.5}$ exposure estimation, we anticipate that the statistical modeling methods can provide more accurate CO exposure estimates than CTMs in regions with extensive ground data. We thus aim to extend our previous work on $PM_{2.5}$, supported by AWS Cloud Computing as well, to CO exposure modeling.

Project Goals and Phases

- *Phase 1*: Due in Winter 2021 – Compiling ground, satellite, and auxiliary (meteorology and land-use) data for CO model development. Validating the statistical model in data-rich modeling domain of North America (Case #1 in Figure 1).
- *Phase 2*: Due in Summer 2022 – 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 Autumn 2022 – Developing a global module by combining continental modules (Case #3 in Figure 1). Predictions between continental boundaries will be driven by both satellite and CTM scaled concentrations with calibration parameters obtained by spatially-interpolating those from neighboring regions.

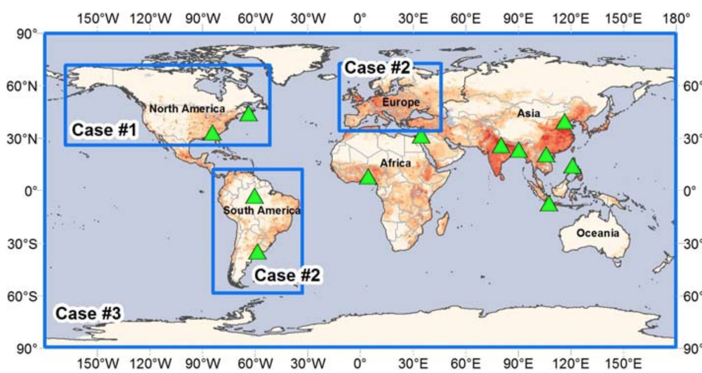

Figure 1: Modeling domains for Cases #1 to #3 (blue boxes).

Experimentation Requirements: AWS Cloud Computing

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

The cloud computing resources we will use include a single EC2 node (m6g.8xlarge, vCPUs = 32, Memory = 128 GiB) with 1 TB storage. We expect to use the EC2 node on a demand basis with 20% utilization per month. The total cost estimate for 12 months is approximately USD 3300. Please see this link regarding the detailed cost estimates.

For this one-year proposal, we aim to apply for AWS research credit of USD 3000 to cover the cost associated with the aforementioned AWS computing resources. We will not reapply for the research credit over the one-year project period.

Additional research cost, including data request and processing, device deployment and maintenance, personnel, and any extra cloud computing cost, will be covered by the Multi-Angle Imager for Aerosols (MAIA) science team and the Multi-angle Imaging SpectroRadiometer (MISR) science team, both at the Jet Propulsion Laboratory (JPL), California Institute of Technology (Subcontract # 1588347 and 1363692).

Summary

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

References

[1] Cohen, A. J., et al. (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." Lancet 389(10082): 1907-1918.

[2] Inness, A., et al. (2019). "The CAMS reanalysis of atmospheric composition." Atmospheric Chemistry and Physics 19(6): 3515-3556.

[3] Inness, A., et al. (2013). "The MACC reanalysis: an 8 yr data set of atmospheric composition." Atmospheric Chemistry and Physics 13(8): 4073-4109.