New Directions: Satellite driven PM$_{2.5}$ exposure models to support targeted particle pollution health effects research

Various adverse health impacts due to exposure to fine particulate matter (PM$_{2.5}$, airborne particles with aerodynamic diameter $<2.5$ μm) have been reported in numerous epidemiological studies (Pope and Dockery, 2006). The World Health Organization estimates that outdoor air pollution (primarily PM$_{10}$ and PM$_{2.5}$) accounts for 800,000 deaths each year, and the burden predominantly occurs in developing countries. Most of exposure-response associations were derived from the United States and a handful of European countries where PM$_{2.5}$ is routinely monitored. Lack of exposure estimates has been a serious limiting factor to evaluate the short- and long-term health outcomes of fine particle pollution in the developing world. In order to identify the health effects due to PM$_{2.5}$ exposure among various confounders, epidemiologists often need to follow a large cohort of people for several years or longer. Assignment of individual exposure levels would require a large and expensive monitoring network. Even if such networks can be established in the near future, relying solely on their observations will likely delay the onset of these studies for years.

With the rapid development of satellite aerosol remote sensing technology, researchers started to explore the possibility of converting satellite-retrieved aerosol optical and microphysical properties to PM$_{2.5}$ mass concentration. So far the most widely used satellite parameter is the aerosol optical depth (AOD). Because the relationship between PM$_{2.5}$ and AOD varies with particle composition, vertical distribution, and size distribution, these factors need to be accounted for in order to develop a robust PM$_{2.5}$ estimation model (Hoff and Christopher, 2008). This can be done in a statistical model with effect modifiers such as meteorology and land use variables (Liu et al., 2005). This emerging technique is particularly valuable for developing countries to conduct PM$_{2.5}$ health effects research because it can generate long-term PM$_{2.5}$ exposure estimates over large regions. For example, the Moderate Resolution Imaging Spectroradiometers (MODIS) aboard U.S. National Aeronautics and Space Administration (NASA)'s Terra and Aqua satellites achieve global coverage of AOD measurements every 1–2 days at 10 km nominal spatial resolution. MODIS data is available from Terra since early 2000 and from Aqua since 2002. The purpose of this article is to provide thoughts on how to develop such a PM$_{2.5}$ exposure modeling system using MODIS data as an example in areas with very sparse or no ground monitoring networks.

The goal of a satellite-based PM$_{2.5}$ model is to generate a spatially resolved, retrospective database to support research on the acute and chronic effects of PM$_{2.5}$ exposure. This process involves five key steps as illustrated by Fig. 1. The study region is first defined by the availability of geocoded population health records. The availability of ancillary data such as meteorology and land use must also be taken into account. Furthermore, the study region should be large enough to allow significant spatial contrast of PM$_{2.5}$ levels, while it must remain logistically manageable by the field team conducting the ground-level PM$_{2.5}$ measurements.

Most current statistical PM$_{2.5}$ models rely on ground observations to calibrate satellite retrievals. In areas without any fixed PM$_{2.5}$ monitoring sites, field sampling must be conducted with portable devices. Properly weather-proofed portable filter-based PM$_{2.5}$ samplers allow field teams to set up a temporary network with minimum requirements for infrastructure. Electricity is often all that is needed in the field. In remote locations, samplers can be powered by rechargeable batteries or solar panels, and portable GPS devices can be used to record site coordinates which are often used as covariates in the model. Sampling locations should reflect PM$_{2.5}$ spatial heterogeneity at different spatial scales under typical weather conditions. In areas which lack ground observations, such knowledge may be gained through satellite observations and atmospheric chemistry models that do not rely on ground truth to predict PM$_{2.5}$ levels (Liu et al., 2004; van Donkelaar et al., 2010). Given the 10 km spatial resolution of MODIS AOD data, the majority of the portable samplers should be placed at least 10–20 km apart. The author’s personal experience while participating in NASA's DISCOVER-AQ campaign in Baltimore, MD in July 2011 suggests that a dedicated field technician was able to manage 5–6 sites at this distance in urban environment. Therefore, a team of four field technicians is able to maintain a network of 20–24 monitors covering a large metropolitan area (approximately 50 km × 50 km).

At least 12 months of sampling is needed for the statistical model to capture seasonal variation of PM$_{2.5}$ levels. The required total number of observation-days is dependent on the complexity of the model. For example, Lee et al. (2011) developed a linear mixed effects (LME) model in Massachusetts U.S. with 576 PM$_{2.5}$ – MODIS AOD observations in 2003 from 26 ground monitors (Lee et al., 2011). Assuming 40–50% missingness due to cloud cover, 60 sampling days (2–3 weeks per season) are necessary for a network of 20 monitors to generate 600 PM$_{2.5}$ – MODIS AOD observations in a year. Satellite instruments with less temporal coverage will require additional sampling days.

Meteorological parameters such as temperature, relative humidity, and wind velocity influence the size, composition and mixing of particles, therefore are important modifiers of the PM$_{2.5}$ – AOD relationship. These common parameters are available worldwide from over 10,000 surface weather stations affiliated with the World Meteorological Organization (WMO). In addition, the U.S. National Oceanic and Atmospheric Administration (NOAA), NASA and the European Centre for Medium-Range Weather Forecasts (ECMWF) all have generated global reanalysis datasets at

http://dx.doi.org/10.1016/j.atmosenv.2012.11.043
coarser spatial resolutions (~1 km). Land use parameters such as urban areas, major road ways, and forest cover are directly linked to emission sources, therefore are valuable ancillary variables in the satellite-based models. Freely available global datasets are available although they do not offer the most detailed land cover classifications. For example, the U.S. Geological Survey's (USGS) Land Cover Institute (LCI) hosts the 1-km resolution Global Land Cover Characteristics Database jointly developed by the U.S. and European partners.

Multiple statistical modeling frameworks have been proposed in recent years including multiple linear regression, LME model, generalized additive model (GAMs), generalized additive mixed model (GAMM), and Bayesian hierarchical model (Hu et al., 2012; Lee et al., 2011; Liu et al., 2009). With multiple years of ground PM_{2.5} observations in the U.S., these advanced modeling systems are able to predict daily PM_{2.5} concentrations with ~25% relative error. It is beyond the scope of this article to evaluate these models. Nonetheless, the more complex models generally have more robust performance but they also require more extensive data support. It should be noted that none of these models have been tested in developing countries with much higher PM_{2.5} levels and different emission source profiles. Given the limited ground data support, it would be appropriate to start by building a simpler model, then cross-validate and enhance it as more PM_{2.5} measurements become available. The final step is to use the statistical models to hindcast historical PM_{2.5} concentrations on a modeling grid during the entire satellite mission as long as all the predictor variables are complete. The spatial resolution of this grid should be comparable with satellite data resolution. This mission-wide PM_{2.5} database can be updated as new versions of satellite data products are released, more ground measurements cumulate and the model itself improves. It is worth noting here that the models that use daily calibration of the PM_{2.5} – AOD relationship need to be modified to rely on the daily relationship obtained during the sampling period.

In conclusion, statistical models driven by satellite aerosol remote sensing data are valuable tools to generate decade-long PM_{2.5} exposure estimates over a large geographical region. Compared with atmospheric chemistry models which rely heavily on detailed emissions inventory to generate accurate PM_{2.5} estimates, they require significantly less data support and computing resources. An added benefit is that statistical models can provide uncertainty estimates of predicted PM_{2.5} concentrations (Liu et al., 2005). With limited ground observations, the spatial and temporal characteristics of regional PM_{2.5} levels can be established quantitatively in a relatively short period. This ability should greatly speed up the progress of air pollution health effects studies in much needed developing countries.

References


Yang Liu

Department of Environmental Health, Emory University, Rollins School of Public Health, Atlanta, GA 30322, USA

E-mail address: yang.liu@emory.edu