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

Ying Liu $^1$ and Guofeng Cao $^2$

AutoCarto, November 2022

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

$^2$Department of Geography
University of Colorado Boulder
guofeng.cao@colorado.edu
https://www.spatstat.com
Motivation

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)
Motivation

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

- Typical studies have been performed based on the limited ground-based PM$_{2.5}$ monitoring observations or coarse resolution PM$_{2.5}$ images;
- Most of the studies focus on short-term effects of 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 PM$_{2.5}$ concentration maps;
- to explore the long-term PM$_{2.5}$-morbidity and disparities in the state of Arizona with 10 years of hospital admission records
High-resolution mapping of PM$_{2.5}$

- **Statistical models**
  - Build statistical relationship between 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

- **Geophyiscal models**
  - Model the diffusion of chemical compositions in atmospheric dynamics
  - e.g., GEOS-Chem chemical transport model to convert the AOD components to PM$_{2.5}$
  - **Advantages:** not directly rely on 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)
High-resolution mapping of $PM_{2.5}$

Machine learning-based regression kriging for downscaling:

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

where $\epsilon(s) \sim GP(\mu, \Sigma)$

- **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:

$$\hat{z}_{1km}^{1km} = F_{RF}(z_{10km}^{1km}, x_{ntl}^{1km}, x_{ndvi}^{1km}, x_{ele}^{1km}; \widehat{\beta}) + \epsilon$$

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

<table>
<thead>
<tr>
<th></th>
<th>( R^2 )</th>
<th>ME (µg/m(^3))</th>
<th>MAE (µg/m(^3))</th>
<th>RMSE (µg/m(^3))</th>
<th>t-test</th>
<th>F-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coarse-PM(_{2.5})</td>
<td>0.533</td>
<td>2.540</td>
<td>2.749</td>
<td>3.522</td>
<td>0.000***</td>
<td>0.000***</td>
</tr>
<tr>
<td>GWR-refined-PM(_{2.5})</td>
<td>0.600</td>
<td>0.215</td>
<td>1.732</td>
<td>2.299</td>
<td>0.000***</td>
<td>0.000***</td>
</tr>
<tr>
<td>RFRK</td>
<td>0.720</td>
<td>−0.046</td>
<td>1.309</td>
<td>1.836</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>
High-resolution mapping of $PM_{2.5}$

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

![Map images showing PM$_{2.5}$ distribution with GEOS, GWR, and RFRK methods comparing Atlanta and Athens, GA regions.](image)
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
Data Preprocessing

**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

\[ RR^d(c) = \frac{EP^d_c / (EP^d_c + EN^d_c)}{CP^d_{c0} / (CP^d_{c0} + CN^d_{c0})} \]

- \( RR^d(c) \): PM\textsubscript{2.5}-Morbidity RR under PM\textsubscript{2.5} concentration \( c \)
- \( EP^d_c (EN^d_c) \): number of persons with positive (negative) outcomes in exposure group
- \( CP^d_{c0} (CN^d_{c0}) \): number of persons with positive (negative) outcomes in control group
- \( c_0 \): theoretical-minimum-risk PM\textsubscript{2.5} level
An iterative way to find $c_0$:

$$RR^d(c_0) = \frac{EP^d_{c_0}}{CP^d_{c_0} / (CP^d_{c_0} + CN^d_{c_0})},$$

where $c'_0 = c_0 + 0.1$

$c_0$ for diseases of interests:

<table>
<thead>
<tr>
<th>Disease</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-cause</td>
<td>3.3 $\mu$g/m$^3$</td>
</tr>
<tr>
<td>Skin cancer</td>
<td>3.8 $\mu$g/m$^3$</td>
</tr>
<tr>
<td>Asthma</td>
<td>3.7 $\mu$g/m$^3$</td>
</tr>
<tr>
<td>Heart diseases</td>
<td>3.5 $\mu$g/m$^3$</td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td>3.6 $\mu$g/m$^3$</td>
</tr>
<tr>
<td>Cerebrovascular diseases</td>
<td>3.6 $\mu$g/m$^3$</td>
</tr>
</tbody>
</table>
Integrated exposure response

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

- \( \alpha, \beta, \) and \( \delta \) 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 $\text{PM}_{2.5}$-morbidity across Subpopulation Groups
• 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 $\text{PM}_{2.5}$ 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 ...

