Air Pollution and Gastrointestinal Diseases in Utah

Melissa May Maestas, *almost* Ph.D.

Research Associate
Family Medicine
Michigan State University

PhD Candidate
Atmospheric Sciences
University of Utah

Measurements, Modeling, and Data Integration for Air Quality and Health
September 21, 2016
Reducing Air Pollution in Detroit Intervention Study (RAPIDS)

Masako Morishita, Robert Brook, Rosemary Ziemba, Sara Adar, and Cathie Spino

- Volunteers: residents of 55+ community in downtown Detroit
- Various filters on home air purifiers
- Concurrent personal, indoor, and outdoor PM$_{2.5}$ monitoring
- Cardiovascular measurements

More on RAPIDS

Photo
https://taubmancollege.umich.edu/events/2015/09/11/inaugural-fall-lecture-detroit-city-planner-maurice-cox
Does air pollution affect Gastrointestinal (GI) Disease?

Both photos taken on January 7, 2011 along Grandeur Peak Ridgeline Trail South of Parley’s Canyon/I-80

11:33 am

4:38 pm

Photos taken by Professor Dave Bowling as part of the Persistent Cold Air Pool Study (PCAPS)

pcaps.utah.edu/data/PHOTOS/browser
Utah Study Acknowledgements

- Committee: Kevin Perry, Court Strong, John Lin, Kathryn Peterson, Ken Smith
- Liz Joy, Kyle John Morgan, Kent Korgenski, Jacob Wilkes, Kerry Rowe, Amanda Borba (Intermountain Healthcare)
- Heidi Hansen, Alison Fraser, Diana Lane, Yuan Wan (UPDB, Demography)
- Rafael Firszt (Pediatric Immunology)
- Micky Daurelle (UU EDW)
- Josi Wood (UU IRB)
- Jahn Barlow, Marybeth Janerich (UU RGE)
- Phoebe Mcneally (DigitLab)
- Center for High Performance Computing (CHPC) Staff
- Funding through a University of Utah seed grant from the Vice President of Research
O$_3$ monitors
• Centroids of census block groups
Kriging – Variogram Fitting

- Take a weighted average of the available data from monitors to estimate pollution concentration at a location without a monitor
- Fit a Variogram Model to determine the weights in the weighted average
- Variogram fit by
  - Pollutant
  - Airshed
  - Season (all-year, cool/warm, 4-seasons)
Finding Best Fits by Pollutant, Airshed, & Season

- Tested 3 models: Gaussian, Exponential, Spherical
- Explored parameter space to back out appropriate model parameters
- Calculate Residuals and use to test quality of fit:
  - Mean
  - Variance
  - Normality
  - Correlation
- Lowest RMSE

Kitanidis, 1997, section 3.4
Less agreement between monitors

More agreement between monitors

Footprint: representative distance over which each monitor is relevant

Pollution concentrations become decoupled at this distance

Calculated from fit parameters

Far apart: concentrations are decoupled

Close together: similar concentration

Close together: High correlation

Far apart: Low correlation
Some fits didn’t work out as well…

Correlation increases at further distances?!!

Wasatch Front NO$_2$, JJA
Quality of fit varies by pollutant

- # and spatial distribution of monitors
- Emissions patterns
- Chemistry
- Topography
## Wasatch Front Footprints

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>Season</th>
<th>Footprint (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM$_{2.5}$</td>
<td>DJF</td>
<td>16.50</td>
</tr>
<tr>
<td></td>
<td>MAM</td>
<td>24.12</td>
</tr>
<tr>
<td></td>
<td>JJA</td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td>SON</td>
<td>24.00</td>
</tr>
<tr>
<td>O$_3$</td>
<td>All-year</td>
<td>34.74</td>
</tr>
</tbody>
</table>
Is the effort of fitting and kriging worthwhile?

- Is it an improvement over the simpler methods?
- Leave-one-out cross validation
  - Leave one monitor out and use others to “predict” concentrations of “left out” monitor
- Analyze difference between “predicted” and observed
4 options for matching data from monitor to zip code

- County average
- Closest monitor
- Closest monitor within footprint (CFP)
  - Don’t assign pollution data if no monitor within footprint
- Kriging (weighted average, interpolation)
Cross Validation: Leave-one-out

\[ \delta_k = z(x_k) - \hat{z}_k \]

Applied to all 4 methods for comparison
Is the effort of fitting and kriging worthwhile?

Bars left of dashed lines: variogram fitting was an improvement over using closest or county average

CFP = closest monitor, limited to footprint found from kriging fit
Data Elements:
• Date of disease event
• Gender
• Age
• Zip code of residence

Inflammatory Bowel Disease (IBD) data merged with pollution data using 4 separate estimation methods
Analysis Method: Case-crossover design (CCOD)

- “Event” is ED visit or Prednisone Rx, etc.
- All individuals have experienced event of interest
  - Everyone is their own control
- Control periods represent “the usual levels of exposure”
- Hazard period represents “exposure of interest” (4-day, 1-week)

Jaakkola, 2003
CCOD

- Uses Logistic Regression
- Regressed together: $\text{PM}_{2.5}$, $\text{O}_3$, $\text{NO}_2$, Temperature, and Medical Data
- Results are given as Odds Ratio (OR), Confidence Interval (CI), and p-value
- In this analysis:
  - OR > 1: air pollution would be detrimental
  - OR < 1: air pollution would be protective
### Method 4D: Symmetric Bidirectional 4 day window

<table>
<thead>
<tr>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June 1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>16 (1)</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>23 (2)</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>29</td>
<td>30 (3)</td>
<td>July 1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7 (event)</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14 (4)</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>21 (5)</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>28 (6)</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>August 1</td>
</tr>
</tbody>
</table>

Use mean & max (separate regressions)
Any events that occur within 56 days of the event are disregarded
“event” is disease flare, i.e., Prednisone Rx
Method 1W: Symmetric Bidirectional ± 1 week

<table>
<thead>
<tr>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June 1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24 control 1</td>
<td>25</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>29</td>
<td>30</td>
<td>July 1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7 (event)</td>
<td>8 control 2</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>August 1</td>
</tr>
</tbody>
</table>

Use mean & max (separate regressions)
Any events that occur within 56 days of the event are disregarded
“event” is disease flare, i.e., Prednisone Rx
## IBD: Prednisone Prescription

- **Grey:** non-significant (p-value $\geq 0.10$)
- **Black:** $p$-value < $0.10$ w/ OR > 1
- **+** $p$-value < $0.05$ w/ OR > 1
- ***** $p$-value < $0.01$ w/ OR > 1

<table>
<thead>
<tr>
<th>PM$_{2.5}$</th>
<th>Krig</th>
<th>CFP</th>
<th>Closest</th>
<th>County Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>75 CU mean</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>*+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>100 CU mean</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>*+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>125 CU mean</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>*+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Contin. mean</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Contin. max</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>75 CU max</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>100 CU max</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>125 CU max</strong></td>
<td>4D</td>
<td>1W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4D</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IBD: Prednisone Prescription

Dose-response
Conclusions

- Kriging and CFP methods generally improved estimation over closest monitor and county average.
- Closest monitor without a restricted footprint almost never performed better than other estimation methods.
- Possible association between prednisone prescriptions and PM$_{2.5}$ among IBD patients.
- First study looking at acute effects of air pollution on IBD.
- Policy implications:
  - GI patients as ‘vulnerable population’ for air pollution.
THANKS!

Questions?
**IBD: GI-tract infections**

- **Grey**: non-significant (p-value $\geq 0.10$)
- **Black**: p-value $< 0.10$ w/ OR $>1$
- **+**: p-value $< 0.05$ w/ OR $>1$
- *****: p-value $< 0.01$ w/ OR $>1$

<table>
<thead>
<tr>
<th>PM$_{2.5}$</th>
<th>Krig</th>
<th>CFP</th>
<th>Closest</th>
<th>County Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>U</td>
<td>C</td>
<td>I</td>
</tr>
<tr>
<td>75 CU mean</td>
<td>4D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 CU mean</td>
<td>4D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125 CU mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contin. mean</td>
<td>4D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contin. max</td>
<td>4D</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**IBD: C diff infections**

- **Grey**: non-significant (p-value $\geq 0.10$)
- **Black**: p-value $< 0.10$ w/ OR $> 1$
- **+**: p-value $< 0.05$ w/ OR $> 1$
- *****: p-value $< 0.01$ w/ OR $> 1$
EoE & $O_3$

- **Grey**: non-significant (p-value $\geq 0.10$)
- **Black**: p-value $< 0.10$ w/ OR $> 1$
- **+**: p-value $< 0.05$ w/ OR $> 1$
- *****: p-value $< 0.01$ w/ OR $> 1$
EoE & O₃

4-day

A (170) 4De
A (170) 4Dx
D (94) 4Dx
I (77) 4Dx

1-Week

A (171) 1We
A (171) 1Wx
D (94) 1Wx
I (77) 1Wx

O₃ OR

0.98 0.99 1.00 1.01 1.02 1.03 1.04 1.05 1.06 1.07 1.08

0.049
0.028
0.046
0.07
0.029
0.095
0.25
0.25
### EoE & PM$_{2.5}$

- **Grey:** non-significant (p-value $\geq 0.10$)
- **Black:** p-value $< 0.10$ w/ OR $> 1$
- **+** p-value $< 0.05$ w/ OR $> 1$
- ***:** p-value $< 0.01$ w/ OR $> 1$

<table>
<thead>
<tr>
<th>PM$_{2.5}$</th>
<th>Krig</th>
<th>CFP</th>
<th>Closest</th>
<th>County Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>75 CU mean</td>
<td>A</td>
<td>D</td>
<td>I</td>
<td>H</td>
</tr>
<tr>
<td>100 CU mean</td>
<td>A</td>
<td>D</td>
<td>I</td>
<td>H</td>
</tr>
<tr>
<td>125 CU mean</td>
<td>A</td>
<td>D</td>
<td>I</td>
<td>H</td>
</tr>
<tr>
<td>Contin. mean</td>
<td>A</td>
<td>D</td>
<td>I</td>
<td>H</td>
</tr>
<tr>
<td>Contin. max</td>
<td>A</td>
<td>D</td>
<td>I</td>
<td>H</td>
</tr>
</tbody>
</table>

**Legend:**
- A: All
- D: Daily
- I: Interquartile
- H: Hourly